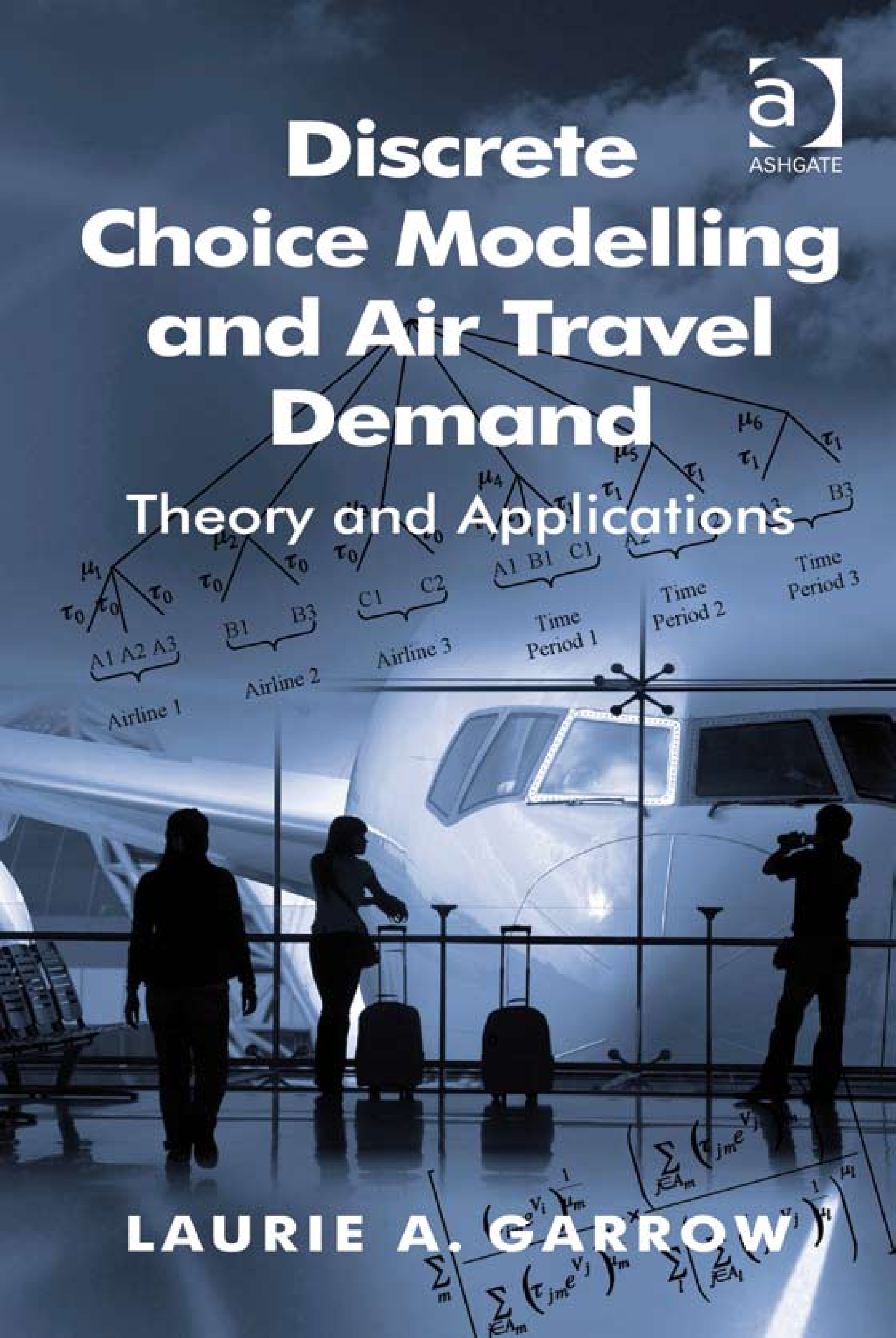
Discrete Choice Modelling and Air Travel Demand

Theory and Applications



LAURIE A. GARROW

ASHGATE

DISCRETE CHOICE MODELLING AND AIR TRAVEL DEMAND

To my parents, Bob and Laura Bowler,
who instilled in me a love of math and a passion for writing.
I dedicate this book to them, as they celebrate 40 years of marriage together this year.

And to my husband, Mike,
who has continuously supported me and encouraged me to pursue my dreams.

Discrete Choice Modelling and Air Travel Demand

Theory and Applications

LAURIE A. GARROW
Georgia Institute of Technology, USA

ASHGATE

Published by
Ashgate Publishing Limited
Wey Court East
Union Road
Farnham
Surrey, GU9 7PT
England

Ashgate Publishing Company
Suite 420
101 Cherry Street
Burlington
VT 05401-4405
USA

www.ashgate.com

British Library Cataloguing in Publication Data
Garrow, Laurie A.
Discrete choice modelling and air travel demand : theory and applications.
1. Air travel--Mathematical models. 2. Aeronautics, Commercial--Passenger traffic--Mathematical models.
3. Choice of transportation--Mathematical models.
I. Title
387.7'015118-dc22

ISBN: 978-0-7546-7051-3 (hbk)
978-0-7546-8126-7 (ebk) V

Library of Congress Cataloging-in-Publication Data
Garrow, Laurie A.
Discrete choice modelling and air travel demand : theory and applications / by Laurie A. Garrow.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-7546-7051-3 (hardback) -- ISBN 978-0-7546-8126-7 (ebook)
1. Aeronautics, Commercial--Passenger traffic--Mathematical models. 2. Scheduling --Mathematics. 3. Demand (Economic theory)--Mathematical models. 4. Discrete-time systems. I. Title.
HE9778.G37 2009
387.7'42011--dc22

2009031152

Mixed Sources
Product group from well-managed forests and other controlled sources
www.fsc.org Cert no. SA-COC-1565
© 1996 Forest Stewardship Council
FSC

Printed and bound in Great Britain by
MPG Books Group, UK

Contents

This page has been left blank intentionally

List of Figures

List of Tables

List of Abbreviations

ARC	Airlines Reporting Corporation
ASC	alternative specific constant
BSP	Billing and Settlement Plan
BTS	Bureau of Transportation Statistics
DB1A	Origin and Destination Data Bank 1A (US DOT data)
DB1B	Origin and Destination Data Bank 1B (US DOT data)
CDF	cumulative distribution function
CRS	computer reservation system
ESML	exogenous sampling maximum likelihood
GEV	generalized extreme value
GNL	generalized nested logit
HeNGEV	heterogeneous covariance network generalized extreme value
HEV	heteroscedastic extreme value
IATA	International Air Transport Association
IIA	independence of irrelevant alternatives
IID	independently and identically distributed
IIN	independence of irrelevant nests
IPR	interactive pricing response
LL	log likelihood
MIDT	Marketing Information Data Tapes
ML	maximum likelihood
MNL	multinomial logit
MNP	multinomial probit
MPO	metropolitan planning organization
NetGEV	network generalized extreme value
NL	nested logit
N-WNL	nested-weighted nested logit
OAG	Official Airline Guide
OGEV	ordered generalized extreme value
OGEV-NL	ordered generalized extreme value-nested logit
OR	operations research
PCL	paired combinatorial logit
PD	product differentiation
PDF	probability density function
PNR	passenger name record
QSI	quality of service index
RM	revenue management
SL	simulated likelihood

SLL	simulated log likelihood
US DOT	United States Department of Transportation
WESML	weighted exogenous sampling maximum likelihood
WNL	weighted nested logit

List of Contributors

Gregory M. Coldren is President of Coldren Choice Consulting Ltd. where he develops logit-based demand forecasting models. He also teaches part-time at several colleges in Maryland and Pennsylvania. He received a Ph.D. in Civil and Environmental Engineering in 2005 from Northwestern University, and is the lead author of several publications.

Frank S. Koppelman is Founding Principal of Midwest System Sciences, Inc., Managing Partner of ELM-Works, LLC, and Professor Emeritus of Civil and Environmental Engineering at Northwestern University. He is an expert in the development, application and interpretation of advanced discrete choice models, travel behavior analysis methods, and consumer choice modeling for public and private firms.

Misuk Lee received her Ph.D. in Industrial and Systems Engineering at the Georgia Institute of Technology. She holds B.S. and M.S. degrees from Seoul National University (2000, 2002). She is currently doing post-doctoral research at the Georgia Institute of Technology with Laurie Garrow and Mark Ferguson. Her research interests include stochastic processes, discrete choice models, transportation systems, and consumer behavior.

Jeffrey P. Newman received his Ph.D. in Civil and Environmental Engineering in 2008 from Northwestern University and is a senior partner of ELM-Works, LLC. He has worked with Michel Bierlaire at the École Polytechnique Fédérale de Lausanne, and is currently doing post-doctoral research at the Georgia Institute of Technology with Laurie Garrow and Mark Ferguson.

This page has been left blank intentionally

Acknowledgements

I have always been a firm believer that an individual's success is not possible without the support and backing of family, friends, and colleagues. The completion of this book is no different, and I am indeed indebted to many individuals who helped to make this book become a reality.

I would first like to acknowledge Roger Parker for encouraging me to write this book and for providing me with valuable feedback on initial chapter outlines. I always look forward to our heated debates on the "best" way to model customer behavior. I also owe Roger a note of appreciation for encouraging me to present our work at the 2005 Air Transport Research Society meeting in Rio de Janeiro, which is where we met Guy Loft of Ashgate Publishing and jointly conceived the vision for this book.

I am also very grateful to my Georgia Tech colleague, Mike Meyer, for the tremendous support he has—and continues—to provide. Mike has been an invaluable mentor and has provided me with leadership opportunities and guidance that have dramatically influenced the professor and writer I am today. I feel particularly honored that I have been able to follow in his footsteps and complete a textbook while an Assistant Professor—"just like he did" with Eric Miller when they were first starting out their academic careers.

The completion of this book would also not be possible without the support of Frank Koppelman, Emeritus Professor of Civil Engineering at Northwestern University. Frank was my doctoral advisor who was always very supportive of my dream to work for an airline while pursuing my graduate studies. The years I spent at United learning about revenue management, scheduling, pricing, and operations were some of the most exciting and influential years of my graduate program, and I am indebted to many individuals (too lengthy to list) from United and their Star Alliance Partners who taught me about the airline industry. A large number of my ideas related to how discrete choice models can be leveraged for airline applications were formed during the period I worked at United and was pursuing my doctoral degree under Frank. Many of these ideas were dramatically shaped and refined through interactions with Frank and two of his other doctoral students: Greg Coldren and Jeff Newman. After graduating from Northwestern, I had the opportunity to continue to work with Frank as we developed training courses in discrete choice modeling for the Hellenic Institute of Transport and the San Francisco Metropolitan Transportation Commission. I am very grateful to Frank for allowing me to use material from these training courses, which drew heavily from material in his graduate courses. It has been a delight working with Frank, Greg, and Jeff, and I look forward to many more years of working with them.

In my current role at Georgia Tech, I have also been fortunate to have worked with several colleagues who have helped me gain a better appreciation for the subtleties of how discrete choice models are applied in different disciplines. Key among these colleagues are Marco Castillo, Mark Ferguson, and Pinar Keskinocak.

Another group of individuals who must be acknowledged are my students. Earlier drafts of the text were used in my graduate travel demand analysis classes, and I benefited dramatically from the comments these students provided. I am also particularly grateful to my post-doctoral student, Misuk Lee, who helped derive elasticity formulas provided in Chapter 4 and who was instrumental in helping solve formatting problems I encountered when producing charts from different software programs. Several other current or former doctoral and post-doctoral students have also contributed to the text by helping to derive and/or check proofs, namely Tudor Bodea, Petru Horvath, Melike Meterelliyoz, and Stacey Mumbower. I am also deeply appreciative for the help of Ana Eisenman, one of our master's students who selfishly dedicated part of her "vacation" preparing proof corrections.

I am also grateful to my colleagues who helped proofread the text and provide critical feedback and suggestions for improvement. These individuals include Greg Coldren, Frank Koppelman, Anne Mercier, Mike Meyer, Jeff Newman, Lisa Rosenstein, and Frank Southworth. I am also deeply appreciative of all of the support that Guy Loft and Gillian Steadman of Ashgate Publishing provided me. Numerous other individuals from industry were also influential in helping me tailor the text to aviation practitioners and students from operations research departments. Key among these contributors are Ross Darrow, Tim Jacobs, Richard Lonsdale, Geoff Murray, Roger Parker, David Post, Richard Ratliff, and Barry Smith, in addition to scores of individuals (too numerous to list) who I met through AGIFORS.

Last, but not least, I owe my family a note of appreciation for their support and encouragement. I am particularly grateful to my husband, Mike, for never complaining about the many hours I had my nose buried in my laptop, and to my father, who would diligently call me every week "just to see how the book was coming along."

Preface

I vividly remember the summer day back in 1998 when I left my studio apartment in downtown Chicago, walked to the Clark and Division CTA station, and started the 22-mile journey out to the suburb of Elk Grove Village for my first day as an intern in United Airlines' revenue management research and development group. I had just completed the first year of my doctoral program at Northwestern University under the guidance of Frank Koppelman, an expert in discrete choice models and travel demand modeling. At the same time I was starting my internship with United, Matt Schrag (now Director of Information Technology) was departing for Minneapolis to work for Northwest Airlines. I was presented with the opportunity to work on one of Matt's projects investigating customer price elasticity. The project fit well with my academic background, and I soon found myself heavily engaged with colleagues from Star Alliance Partners collaborating on the project as well as senior consultants; these individuals include Paul Campbell (now Vice-President of Sales at QL2), Hugh Dunleavy (now Executive Vice-President of Commercial Distribution at Westjet), Dick Niggley (now Vice-Chairman of Revenue Analytics), and independent consultants Ren Curry and Craig Hopperstad who had played instrumental roles in developing some of the first airline revenue management and scheduling applications. I could not have asked for a better group of colleagues to introduce me to the airline industry.

At the end of the summer, I continued to work for United and, over the course of the next four years, became involved in a variety of different projects. During this period, I began advocating the use of discrete choice models for different forecasting applications. I have to admit, at the early stages of these discussions, I remember the large number of "off the wall" questions I received from my colleagues. With time, I came to understand and appreciate the underlying motivations for why my colleagues (who had backgrounds in operations research) were asking me these questions. Many of the questions arose due to subtle—yet critically important—differences related to the approaches operations research analysts and discrete choice analysts use to solve problems. For example, while it is natural (and indeed, often a source of pride) for operations research analysts to think in terms of quickly optimizing a problem with thousands (if not millions) of decision variables, it is natural for a discrete choice analyst to first design a sampling plan that decreases model estimation times without sacrificing the ability to recover consistent parameter estimates.

The key objectives, themes, and presentation of this text have been dramatically shaped by these personal experiences. The primary objective of this text is to provide a comprehensive, introductory-level overview of discrete choice models.

The text synthesizes discrete choice modeling developments that researchers and students with operations research (OR) and/or travel demand modeling backgrounds venturing into discrete choice modeling of air travel behavior will find most relevant. In addition, given the strong mathematical background of OR researchers and airline practitioners, a set of appendices containing detailed derivations is included at the end of several chapters. These derivations, frequently omitted or condensed in other discrete choice modeling texts, provide a foundation for readers interested in creating their own discrete choice models and deriving the properties of their models.

In this context, this book complements seminal texts in discrete choice modeling that appeared in the mid-1980s, namely those of Ben-Akiva and Lerman (1985) and Train (1986; 1993). Given that the focus of this text is on applications of discrete choice models to the airline industry, material typically covered in travel demand analysis courses related to stated preference data (such as survey design methods and strategies to combine revealed preference and stated preference data) is not presented. Readers interested in these areas are referred to Louviere, Hensher, and Swait (2000). Additional references that cover a broader range of travel demand modeling methods as well as advanced topics include those by Greene (2007), Greene and Hensher (2010), Hensher, Greene, and Rose (2005), and Long (1997).

The book contains a total of eight chapters. Chapter 1 highlights the different perspectives and priorities between the aviation and urban travel demand fields, which led to different demand modeling approaches. Given that many discrete choice modeling advancements were concentrated in the urban travel demand area, the comparison of major differences between the two fields provides a useful background context. Chapter 1 also describes data sources that are commonly used by airlines and/or researchers to forecast airline demand.

Chapter 2 covers discrete choice modeling fundamentals and introduces the binary logit and multinomial logit (MNL) models (the most common discrete choice models used in practice). Chapter 3 builds upon these fundamentals by describing how correlation, or increased substitution among alternatives, can be achieved by using a nested logit (NL) model structure that allocates alternatives to non-overlapping nests. An emphasis is placed on precisely defining the nested logit model in the context of utility maximization theory, as there are multiple (and incorrect) definitions and formulations of "nested logit" models used in both the discrete choice modeling field and the airline industry. Unfortunately, these "incorrect" definitions are often the default formulation embedded in off-the-shelf estimation software.

Chapter 4 provides an extensive overview of different discrete choice models that occurred after the appearance of the MNL, NL, and multinomial probit models. This chapter, co-authored with Frank Koppelman and Misuk Lee, draws heavily from book chapters written by Koppelman and Sethi (2000) and Koppelman (2008) contained in the first and second editions of the *Handbook of Transport Modeling*. In contrast to this earlier work, Chapter 4 tailors the discussion of

discrete choice models by highlighting those developments that are relevant, from either a theoretical or practical perspective, to the airline industry. A new approach for using an artificial variance-covariance matrix to visualize “breakdowns” (or “crashes” as coined by Newman in Chapter 5) that occur in models that allocate alternatives to more than one nest is presented; the presence of these breakdowns complicates the ability to calculate correlations among alternatives and often results in the need for identification rules (or normalizations) beyond those associated with the MNL and NL models. Appendix 4.1, compiled by Misuk Lee, contains two reference tables that summarize choice probabilities, general model characteristics, direct-elasticities, and cross-elasticities for a dozen discrete choice models. These tables, which use a common notation across all of the models, provide a useful reference.

Chapter 4 also introduces a framework that is used to classify discrete choice models belonging to the Generalized Extreme Value class that allocate alternatives to more than one nest. Generalized nested logit models include all nested structures that contain two levels whereas Network Generalized Extreme Value (NetGEV) models are more general in that they encompasses all nested structures that contain two or more levels. Chapter 4 presents an overview of some of the first empirical applications of three-level models that allocate alternatives to multiple nests. Interestingly, these empirical applications first appeared in airline itinerary choice models, which were occurring in the early 2000’s at approximately the same time that Andrew Daly and Michel Bierlaire were deriving theoretical properties of the NetGEV model. This is one example of the synergistic relationships emerging between the aviation and discrete choice modeling areas; that is, the need within airline itinerary choice applications to incorporate complex substitution relationships has helped drive interest by the discrete choice modeling community to further investigate the theoretical properties of the NetGEV. Chapter 5, authored by Jeff Newman, summarizes theoretical identification and normalization rules he developed for the NetGEV models as part of his doctoral dissertation, completed in 2008. Additional extensions to the NetGEV model, including a model that allocates alternatives across nests as a function of decision-maker characteristics, are also presented in Chapter 5.

Chapter 6 shifts focus from discrete choice models that have closed-form choice probabilities to the mixed logit model, which requires simulation methods to calculate choice probabilities. In contrast to Kenneth Train’s 2003 seminal text on mixed logit models, Chapter 6 synthesizes recent mixed logit empirical applications within aviation (which have been very limited in the context of using proprietary airline data). Chapter 6 also highlights open research questions related to optimization and identification of the mixed logit model, which will be of particular interest to students reading this text and looking for potential dissertation topics.

The primary goal of Chapter 7 is to illustrate how the mathematical formulas and concepts presented in the earlier chapters translate to a practical modeling exercise. Itinerary share data from a major U.S. airline are used to illustrate

the modeling process, which includes estimating different utility functions and incorporating more flexible substitution patterns across alternatives. Measures of model fit for discrete choice models, as well as statistical tests used to compare different model specifications are presented in this chapter. The utility function and market segmentations for the itinerary choice models contained in this chapter reflect those developed by co-authors Coldren and Koppelman and are illustrative of those used by a major U.S. airline.

Chapter 8 summarizes directions for future research and my opinions on how the OR and discrete choice modeling fields can continue to synergistically drive new theoretical and empirical developments across both fields. One area I am personally quite excited about is the ability to observe, unobtrusively in a revealed preference data context, how airline customers search for information in on-line channels. The ability to capture the dynamics of customers' search and purchase behaviors—both within an online session as well as across multiple sessions—is imminent. In this context, I am reminded of the distinction between static and dynamic traffic assignment methods and the many new behavioral and operational insights that we gained when we incorporated dynamics into the assignment model. From a theoretical perspective, I fully expect the availability of detailed online data within the airline industry to drive new theoretical developments and extensions to dynamic discrete choice models and game theory. I look forward to the next edition of this text that would potentially cover these and other developments I expect to emerge from collaborations between the OR and discrete choice modeling fields. It is my ultimate hope that this text helps bridge the gap between these two fields and that researchers gain a greater appreciation for the seemingly "off the wall" questions that are sure to arise through these collaborations.

Laurie A. Garrow

Chapter 1
Introduction

Introduction and Background Context

In Daniel McFadden's acceptance speech of the Nobel Prize in Economics, he describes how in 1972 he used a multinomial logit model based on approximately 600 responses from individual commuters in the San Francisco Bay Area to forecast ridership for a new BART line (McFadden 2001). This study, typically considered the first application of a discrete choice model in transportation, provided a strong foundation and motivation for urban travel demand researchers to transition from modeling demand using aggregate data to modeling demand as the collection of individuals' choices. These choices varied by socio-demographic and socio-economic characteristics, as well as by attributes of the alternatives available to the individual.

At the same time that McFadden and other researchers were investigating forecasting benefits associated with modeling individual choice behavior to support transit investment decisions, the U.S. airline industry was predicting demand for air travel using Quality of Service (QSI) indices. QSI indices were developed in 1957 and predicted how demand would shift among carriers as a function of flight frequency, level of service (e.g., nonstop, single-connection, double-connection) and equipment type (Civil Aeronautics Board 1970). At the time, the airline industry was regulated, fares and service levels were set by the government, and load factors were about 50 percent (e.g., see Ben-Yosef 2005). Competition was based primarily on marketing promotion and image.

The airline industry changed dramatically in 1978 when it became deregulated and airlines could decide where and when to fly, as well as how much to charge passengers (Airline Deregulation Act 1978). Operations research analysts played a critical role after deregulation, helping to design algorithms and decision-support systems to optimize where and when to fly, subject to minimizing costs associated with assigning pilots and flight attendant crews to each flight while ensuring each plane visited a maintenance station in time for required checks and service. A second milestone event happened in 1985, when American Airlines implemented a revenue management system that offered a limited set of substantially discounted fares with advance purchase restrictions as a way to compete with low fares offered by People's Express Airlines; the strategy worked, and People's Express went out of business shortly thereafter (e.g., see Ben-Yosef 2005). A role for operations research had emerged in the revenue management area, with the primary objective of maximizing revenue (or profit) under uncertain demand forecasts, passenger cancellations, and no shows.

The “birth” of operations research in a deregulated airline industry occurred in an era in which computational power was much more limited than it is today. A major airline faced with optimizing schedules that involved coordinating arrivals and departures for thousands of daily take-offs and landings, assigning tens of thousands of pilots and flight attendants to all of these flights (while ensuring all work rules were adhered to), and keeping track of millions of monthly booking transactions, was clearly facing a different problem context than Daniel McFadden and other travel demand modelers. The latter were making demand predictions to help support investment decisions and evaluation of transportation policies for major metropolitan areas. In this context, the use of discrete choice models to help rank different alternatives and assess short-term and long-term forecast variation across different scenarios was of primary importance to decision-makers.

However, from an airline perspective, it would have been computationally impractical to model the choice of every individual passenger (which would require keeping track of all alternatives considered by passengers). Instead, in the U.S. it was (and still is) common to model market-level itinerary share demand forecasts using ticket information compiled by the U.S. Department of Transportation (Bureau of Transportation Statistics 2009; Data Base Products Inc. 2008) and to use time-series and/or simplistic probability models based on product-level booking or flight-level data to forecast demand for flights, passenger cancellation rates, passenger no show rates, etc.

More than thirty years after deregulation, the airline industry is faced with intense competition and ever-increasing pressures to control costs and generate more revenues. Multiple factors have contributed to the current state of the industry, including the increased use of the Internet as a major distribution channel and the increased market penetration of low cost carriers. It is clear that the Internet has transformed the travel industry. For example, in 2007, approximately 55 million (or one in four) U.S. adults traveled by commercial air and were Internet users (PhoCusWright 2008). As of 2004, more than half of all leisure travel purchases were made online (Aaron 2007). In 2006, more than 365 million U.S. households spent a total of $74.4 billion booking leisure travel online (Harteveldt Johnson Stromberg and Tesch 2006).

The market penetration of low cost carriers has also steadily and dramatically grown since the early 1990’s. For example, in 2004, approximately 25 percent of all passengers in the U.S. flew on low cost carriers, and 11 percent of all passengers in Europe flew on low cost carriers (IBM Consulting Services 2004). Importantly, the majority of low cost carriers in the U.S. use one-way pricing, which results in separate price quotes for the departing and returning portions of a trip. One-way pricing effectively eliminates the ability to segment business and leisure travelers based on a Saturday night stay requirement (i.e., business travelers are less likely to have a trip that involves a Saturday night stay). Combine the use of one-way pricing with the fact that the Internet has increased the transparency of prices for consumers and the result is that today, approximately 60 percent of online leisure

travelers purchase the lowest fare they can find (Harteveldt Wilson and Johnson 2004; PhoCusWright 2004).

Within the operations research community, these and other factors have led to an increasing interest in using discrete choice models to model demand as the collection of individuals' decisions, thereby more accurately capturing *how* individuals are making decisions and trade-offs among carriers, price, level of service, time of day, and other factors. To date, much of the research in using discrete choice models for aviation applications has focused in areas where it has been relatively straightforward to identify the alternatives that individuals consider during the choice process (e.g., airlines have itinerary-generation algorithms that build the set of itineraries or paths between origin-destination pairs). In addition, this research has focused on areas in which it would be relatively easy for airlines to replace an existing module (e.g., a no show forecast) that is part of a much larger decision-support system (e.g., a revenue management system). Itinerary share predictions, customer no show behavior, customer cancellation behavior, and recapture rate modeling all belong to this stream of research (e.g., see Coldren and Koppelman 2005a, 2005b; Coldren Koppelman Kasturirangan and Mukherjee 2003; Garrow and Koppelman 2004a, 2004b; Iliescu Garrow and Parker 2008; Koppelman Coldren and Parker 2008; Ratliff 2006; Ratliff Venkateshwara Narayan and Yellepeddi 2008).

More recently, researchers have also begun to investigate how discrete choice models and passenger-level data can be integrated with optimization models at a systems level. Advancements in computing power combined with the ability to track individual consumers through the booking process have spawned a new era of revenue management (RM), commonly referred to as "choice-based" RM. Conceptually, choice-based RM methods use data that effectively track individuals' purchase decisions, as well as the menus of choices they viewed prior to purchase. That is, in contrast to traditional booking data, on-line shopping data provide a detailed snapshot of the products available for sale at the time an individual was searching for fares, as well as information on whether the search resulted in a purchase (or booking). These data effectively enable firms to replace RM demand models based on probability and time-series models with models grounded in discrete choice theory. To date, several theoretical papers on choice-based RM techniques have appeared in the research community and a few empirical studies based on a limited number of markets and/or departure dates have also been reported (e.g., see Besbes and Zeevi 2006; Bodea Ferguson and Garrow 2009; Bront Mendez-Diaz and Vulcano 2007; Gallego and Sahin 2006; Hu and Gallego 2007; Talluri and van Ryzin 2004; van Ryzin and Liu 2004; van Ryzin and Vulcano 2008a, 2008b; Vulcano van Ryzin and Chaar 2008; Zhang and Cooper 2005).

To summarize, it is clear that the momentum for using discrete choice models to forecast airline demand as the collection of individuals' choices is building, and most importantly, this momentum is building both in the travel demand modeling/discrete choice modeling community as well as in the operations research community.

Primary Objectives of the Text

Although the interest in using discrete choice models for aviation applications is building, there has been limited collaboration between discrete choice modelers and optimization and operations researchers. Part of the challenge is that many operations research departments have provided students with a limited exposure to discrete choice models. This is due in part to the fact that the primary affiliation of most discrete choice modeling experts is not with operations research departments, but rather with transportation engineering, marketing, and/or economics departments. The distinct evolution of the discrete choice modeling and operations research fields has resulted in researchers from these fields having different perspectives, research priorities, and publication outlets.

One of the primary objectives of this text is to help bridge the gap between the discrete choice modeling and operations research communities by providing a comprehensive, introductory-level overview of discrete choice models. This overview synthesizes major developments in the discrete choice modeling field that are relevant to the aviation industry and the challenges this industry is currently facing. An emphasis has been placed on discussing the properties of discrete choice models using terminology that is accessible to both the discrete choice modeling and operations research communities, and complementing these discussions with numerous examples. The discrete choice modeling topics covered in the text (that represent only a small fraction of work that has been developed since the early 1970s), provide a fundamental base of knowledge that analysts will need in order to successfully estimate, interpret, and apply discrete choice models in practice. Consequently, it is envisioned that this text will be useful to aviation practitioners, researchers and graduate students in operations research departments, and researchers and graduate students in travel demand modeling.

Important Distinctions Between Aviation and Urban Travel Demand Studies

Given the different backgrounds and perspectives of aviation operations research analysts and urban travel demand analysts, it is helpful to highlight some of the key distinctions between these two areas.

Objectives of Aviation and Urban Transportation Studies

The overall objectives driving demand forecasting studies conducted for aviation firms and studies conducted for government agencies evaluating transportation alternatives in urban areas tend to be quite distinct. Deregulated airlines, such as those in the U.S. that are private firms and are not owned by governments, are generally focused on maximizing net revenue through attracting new customers and retaining current customers while ensuring safe and efficient operations. Many of the problems investigated by operations research analysts reflect this

strong focus on maintaining safe and efficient operations throughout the airline's network (or system). These problems include building robust network schedules and assigning pilots and flight attendants to aircraft in ways that result in fewer aircraft delays and cancellations and fewer passenger misconnections; assigning aircraft to specific airport gates to ensure transfer passengers have sufficient time to connect to their next flight while considering secondary objectives, such as minimizing the average distance that premium passengers need to walk between a loyalty lounge and the departing gate; scheduling multiple flights into a hub to achieve one or more objectives, such as maximizing passenger connection possibilities, minimizing passenger connection times, and/or flattening peak airport staffing requirements; developing efficient processes to screen baggage and minimize the number of bags that are lost or delayed; creating rules that minimize average boarding time for different aircraft types; developing processes that help airlines quickly recover from irregular operations; overbooking flights to maximize revenue while minimizing the number of voluntary and involuntary denied passengers, etc.

Government agencies, in contrast to airlines, are generally focused on predicting demand for existing and proposed transportation alternatives. A broad range of alternatives may be considered and include infrastructure improvements, operational improvements, new tax fees, credits or other policy instruments, etc. Thus, the primary focus of urban transportation studies is centered on supporting policy analysis, which includes gaining a richer understanding of how individuals, households, employers and other institutions will react to different alternatives. Urban travel demand analyses are also often conducted within a systems-level framework (i.e., examined within the entire urban area), in part to ensure equitable allocation of resources and services across different socio-economic and socio-demographic groups.

Data Characteristics of Aviation and Urban Travel Demand Studies

Given the different objectives of aviation firms and government agencies, it is not surprising that the data used for analysis also differ. Within aviation, the strong operational focus within a relatively large system has resulted in decision-support models based almost exclusively on revealed preference data that contain limited customer information. Revealed preference data capture actual passenger choices under current and prior market conditions. The airline industry is characterized by flexible capacity which results in a large number of observations that tend to vary "naturally" or "randomly" within a market or across different markets. For example, in itinerary share models, frequent schedule changes create "natural" variation in the itineraries available to customers; that is, over the course of a year (or even from month to month), individuals are faced with alternatives that vary by level of service, departure and/or arrival times, connection times, operating carriers, prices, etc. In turn, given the dynamic nature of the airline industry and the need for carriers to identify and respond quickly to changes in competitive

conditions, it is highly desirable to design decision-support models that rely heavily on recently observed revealed preference data.

In addition, due to the large number of flights major carriers manage, any customer information stored in databases tends to be limited to that needed to support operations. For example, from an operations perspective, it is important for gate agents to know how many individuals on an arriving flight need wheelchairs; however, knowing the individual's age, gender, and household income level is irrelevant to the ability of the gate agent to make sure a wheelchair is available for the customer, and is thus not typically collected as part of the booking process. Similarly, although algorithms have been developed to reaccomodate passengers automatically to different flights when their original flight experiences a long delay or cancellation, the prioritization of customers is typically based on prior and current travel information. Archival travel information may include the customer's current status in the airline's frequent flyer program and/or the customer's "value" to the airline that considers both the number of trips the customer has purchased on the carrier as well as how much the customer paid for these trips. Current travel information may include the amount the customer paid for the trip, whether the trip is in a market that has a low flight frequency (resulting in fewer reaccommodation opportunities), and whether the cost of reaccommodating the passenger on a different carrier is high (as in the case for an international itinerary).

In contrast to airline applications with an operations focus, urban travel demand studies rely heavily on socio-economic and socio-demographic information, such as an individual's age, gender, ethnicity, employment status, marital status, number and ages of children in the household, residence ownership status and type (owned or rented; single family home, multi-family residence, etc.), household income, etc. These and other variables (such as the make, model, and age of each automobile owned by the household) are inputs to the travel demand forecasts for an urban area. Conceptually, these models create a simulated population that represents characteristics of the existing population in an urban area. Different transportation alternatives and/or combinations of different transportation alternatives are evaluated by testing how different segments of the population respond, assessing system-level benefits (such as reductions in emissions due to shifting trips from automobile to transit or due to modernizing vehicle fleets over time), and identifying any impacts that are disproportionately allocated across different socio-economic groups.

Urban travel demand studies use a wide range of revealed preference, stated preference data, and combinations of revealed and stated preference data. Revealed preference data sources include observed boarding counts on buses and other modes of transportation, observed screen-line counts (or the number of vehicles passing by a certain "screen-line" in a specified time period), travel survey diaries that ask individuals to record every trip made by members of the household over a short period of time (typically two days), intercept surveys that interview current transit users to collect information about their current trip, etc. From a demand forecasting perspective, the socio-demographic and socio-economic variables that

are inputs to urban travel demand models are available, often at a detailed census tract or census block level, from government agencies. Moreover, for many major infrastructure projects (such as a proposed transit project in the U.S. that requests federal funding support), it is expected that demand forecasts will be based on "recent" customer surveys.

Whereas revealed preference data reflect the actual choices made by individuals under current or previous market conditions, stated preference data are collected via surveys that ask individuals to make hypothetical choices by making trade-offs among the attributes of the choice set (such as time, cost, and reliability measures) determined by the analyst. Stated preference data are particularly useful when investigating customer response to new products or transportation alternatives, or when existing and past market conditions do not exhibit sufficient "natural variation" to allow the analyst to estimate how individuals are making tradeoffs (because the number of distinct trade-off combinations is limited). For example, time-of-day congestion pricing is a relatively new concept that has been implemented in different forms throughout the world. Stated preference surveys designed to investigate how commuters and shippers would potentially change their behavior under different congestion pricing alternatives in a major metropolitan area would be valuable for assessing likely outcomes associated with implementing a similar policy in a new area.

Whereas many aviation studies with an operational focus tend to rely heavily on revealed preference data, stated preference data are also used within the airline industry, albeit primarily in marketing departments where new product designs are of primary interest. For example, Resource Systems Group, Inc., a firm located in Vermont, has been conducting an annual survey of air travelers since 2000. This annual stated preference survey has been supported by a wide variety of airlines and government agencies. Consistent with the use of stated preference data seen in the context of urban travel demand studies, these stated preference surveys have supported a range of new product development studies for airlines (e.g., cabin service amenities, unbundling product strategies, passenger preferences for connection times, etc.) Government agencies have also used this panel to investigate changes in passenger behavior after 9/11. Results from some of these studies can be found in Adler, Falzarano, and Spitz (2005), and Warburg, Bhat, and Adler (2006).

To summarize, although both revealed and stated preference data are used in aviation and urban travel demand studies, aviation studies (particularly those with an operational focus that most operations research analysts investigate) are dominated by revealed preference data that contain limited socio-demographic and socio-economic information.

Other Factors that Influence Estimation and Forecasting Priorities

In addition to different objectives and data sources used by aviation and urban travel demand studies, there are several other factors that influence estimation and

forecasting priorities within these two areas. First, the number of observations used during estimation tends to be much smaller for urban travel demand studies (particularly those based on expensive survey data collection methods) than for aviation studies. Second, given that many urban travel demand studies are used to evaluate infrastructure improvements that have a lifespan of several decades, demand forecasts are produced for current year conditions, as well as ten years, twenty years, and/or thirty years in the future. Demand forecasts are created on an "as needed" basis to support policy and planning analysis, are typically used to help evaluate different alternatives, and are not critical to the day-to-day operations of the government agency (thus, optimizing the speed at which parameter estimates of demand models are solved or decreasing the computational time of producing demand forecasts, although important, is typically not the primary concern of urban travel demand modelers). The ability of analysts to measure forecasting accuracy in this context is not always straightforward, particularly if the policy under evaluation is never implemented.

In contrast, the number of observations used to estimate model parameters in aviation studies is quite large (and in some situations can number in the millions). Importantly, demand forecasts are critical to the day-to-day operations of an airline. For example, in revenue management applications it is not uncommon to produce detailed forecasts (defined for each itinerary, booking class, booking period, and point of sale) on a daily or weekly basis. In scheduling applications, demand forecasts that support mid- to long-range scheduling of flights are often updated on a monthly or quarterly basis. It is also important to recognize that in contrast to many urban transportation studies where the relative ranking of alternatives is important, in airline applications forecasting accuracy is critical, and any improvements tend to translate to millions of dollars of annual incremental revenue for a major carrier. Thus, in revenue management applications, it is not uncommon to include a measure of forecasting variance to capture risk associated with having a demand forecast that is too aggressive (that may lead to high numbers of denied boardings) and risk associated with having a demand forecast that is systematically under-forecasting (that may lead to high numbers of empty seats and lost revenue). It is also not uncommon for airlines to monitor the accuracy of their systems on an ongoing basis, and provide feedback to analysts on how well their adjustments to demand forecasts influence overall forecast accuracy.

One area that is common to both aviation and urban travel demand studies relates to accurately modeling and incorporating competitive substitution patterns. For example, in airline itinerary share prediction, an American Airlines' itinerary departing at 10 AM may compete more with other American Airlines' itineraries departing in mid-morning than with itineraries departing after 5 PM on Southwest Airlines. Similarly, in mode choice studies, the introduction of a new light rail system may draw disproportionately more passengers from existing transit services than from auto modes. Much of the recent research related to discrete choice models was focused on developing methods to incorporate more flexible substitution patterns; these developments form the basis of Chapters 3 to 6 of this

text. In summary, Table 1.1 presents the key distinctions between aviation and urban travel demand studies discussed in this section.

Table 1.1 Comparison of aviation and urban travel demand studies

	Aviation	**Urban Transportation**
Objectives	• Maximize revenue • Safe and efficient operations • Customer attraction and retention	• Policy analysis • Behavioral analysis • Systems-level analysis
Demand Data	• Revealed preference (frequent schedule changes) • Limited socio-demographic information	• Revealed and stated preference • Rich socio-demographic and socio-economic information • Census data
Estimation	• Very large data volumes	• Relatively small data volumes
Forecasting	• Frequent (daily to monthly) • Forecasting accuracy and variability both important	• Driven by policy needs • Forecasts used to provide relative ranking of alternatives
Competition among Alternatives	• Critical	• Critical

Overview of Major Airline Data

Given many students have limited knowledge of and exposure to airline data sources, this section presents a brief overview of some of the most common data used by airlines and/or that are publically available. The data covered in this section are not exhaustive, but are representative of the different types of demand data (bookings and tickets), supply data (schedule), and operations data (check-in, flight delays and cancellations) used in aviation applications.

Booking Data

Booking and ticketing data contain information about a reservation made for a single passenger or a group of passengers travelling together under the same reservation confirmation number, which is often referred to as a passenger name record (PNR) locator. Any changes made to the booking reservation (passenger cancels reservation, passenger requests different departure date and flight, airline moves passenger to a different flight due to schedule changes that occur

pre-departure, etc.) are included in these booking data. The difference between booking and ticketing databases relates to whether the passenger has paid for the reservation. A reservation, or booking request, that has been paid for appears in both booking and ticketing databases, whereas a booking request that has not yet been paid for appears only in a booking database.

Booking databases are maintained by airlines and computer reservation systems (CRS) and are generally not accessible to researchers. Booking data are typically stored at flight and itinerary levels of aggregation and contain information including the passenger's name, PNR locator, booking date, booking class, ticketing method (e.g., electronic or paper ticket), and booking channel (e.g., the airline's website; a third party website such as Travelocity©, Orbitz©, or Expedia©; the airline's central reservation office, etc.) Information about the specific flights or sequence of flights the passenger has booked is also provided, for example, each flight is identified by its origin and destination airports, departure date, flight number, departure and arrival times, and marketing and operating carriers. By definition, a marketing carrier is the airline who sells the ticket whereas the operating carrier is the airline who physically operates the flight. For example, a code-share flight between Delta and Continental could be sold either under a Delta flight number or a Continental flight number. However, only one plane is flown by either Delta or Continental—this is the operating carrier.

Booking databases also contain passenger information required for operations, for example, if the passenger has requested a wheelchair and/or a special meal, is travelling with an infant, is a member of the marketing carrier's frequent flyer program, etc. Note that the price associated with the booking reservation is not always stored with the booking database. Detailed price information for those booking reservations that were actually paid for is contained in ticketing databases.

As noted earlier, airline carriers maintain their own booking databases. However, passengers can make reservations via a variety of different channels. Prior to the increased penetration of the Internet, it was common for passengers to make reservations with travel agents who accessed the reservations systems of multiple airlines via computer reservations systems (CRS) such as Amadeus (2009), Galileo (2009), Sabre (2009), and Worldspan (2009). CRS data (also called Marketing Information Data Tapes (MIDT) data) are commercially available and compiled from several CRSs. In the past, CRS data provided useful market share information. However, Internet bookings and carrier direct bookings (such as those made via the airline's phone reservation system) are not captured in this database, and the reliability and usefulness of this dataset has deteriorated over the last decade.

Lack of prior booking information for a new (often non-U.S.) market is also a challenge, i.e., the lack of revealed preference data in new markets requires airlines to predict demand using stated preference surveys or by using revealed preference data from markets considered similar to the new markets they want to enter. At times, an important behavioral factor can be overlooked. A recent example is the

$25 million investment that SkyEurope made in the airport in Vienna, Austria, to offer low cost service that competes with Austrian Airlines. Originally, SkyEurope planned to capture market share in Vienna using one of the strategies often seen with low cost airlines, i.e., through concentrating service in a secondary airport that was close to Vienna that would be able to draw price-sensitive customers from Vienna. However, in this case, the secondary airport, Bratislava, Slovakia, was in a different country and SkyEurope discovered that passengers were reticent to cross the border separating Austria and Slovakia to travel by air, despite the short driving distance. In light of this customer behavior, SkyEurope made the decision to invest in Vienna in order to capture market share from that city (Karatzas 2009).

Ticketing Data

Ticketing databases are similar to booking databases, but provide information on booking reservations that were paid for. Carriers maintain their own ticketing databases, but there are other ticketing databases, some of which are publically available. One of the most popular ticketing databases used to investigate U.S. markets is the United States Department of Transportation (US DOT) O*rigin and Destination Data Bank 1A or Data Bank 1B* (commonly referred to as DB1A or DB1B). The data are based on a 10 percent sample of flown tickets collected from passengers as they board aircraft operated by U.S. airlines. The data provide demand information on the number of passengers transported between origin-destination pairs, itinerary information (marketing carrier, operating carrier, class of service, etc.), and price information (quarterly fare charged by each airline for an origin-destination pair that is averaged across all classes of service). Whereas the raw DB datasets are commonly used in academic publications (after going though some cleaning to remove frequent flyer fares, travel by airline employees and crew, etc.), airlines generally purchase Superset data from Data Base Products. Superset is a cleaned version of the DB data that is cross-validated against other data-sources to provide a more accurate estimate of the market size. (See the websites for the Bureau of Transportation Statistics (2009) and Data Base Products Inc. (2008) for additional information.) Importantly, the U.S. is one of the few countries that requires a 10 percent ticketing sample and makes this data publically available.

There are two other primary agencies that are ticketing clearinghouses for air carriers. The Airlines Reporting Corporation (ARC) handles the majority of tickets for U.S. carriers and the Billing and Settlement Plan (BSP) handles the majority of non-US based tickets (Airlines Reporting Corporation 2009; International Air Travel Association 2009). In the U.S., data based on the DB tickets differ from the ticketing data obtained from ARC. First, DB data report aggregate information using quarterly averages and passenger counts and ARC data contain information about individual tickets. Second, DB data contain a sample of tickets that were used to board aircraft, or for which airline passengers "show" for their flights. In contrast, ARC data provide information about the ticketing process from the *financial perspective*. Thus, prior information is available for events that trigger

a cash transaction (purchase, exchange, refund), but no information is available for whether and how the individual passenger used the ticket to board an aircraft; this information can only be obtained via linking the ARC data with airlines' day of departure check-in systems. Third, ARC ticketing information does not include changes that passengers make on the day of departure; thus, the refund and exchange rates will tend to be lower than other rates reported by airlines or in the literature. Finally, whereas DB data are publically available, ARC data (in disguised forms to protect the confidentiality of the airlines) are available for purchase from ARC.

Schedule Data

Flight and itinerary schedule data are based on official airline schedules produced by the Official Airline Guide (OAG) (OAG Worldwide Limited 2008). OAG contains leg-based information on the origin, destination, flight number, departure and arrival times, days of operation, leg mileage, flight time, operating airline, and code-share airline (if a code-share leg). It also provides capacity estimates (i.e., the number of itineraries and seats) for each carrier in a market. Garrow (2004) describes how the OAG data, which contain information about individual flights, are processed to create itinerary-level information for representing "typical" service offered by an airline and its competitor. Specifically, Garrow reports on the process used by one major airline as follows: "Monthly reports are created using the flight schedule of one representative week defined as the week beginning the Monday after the ninth of the month. For example, flights operated on Wednesday, March 13, 2002, are used to represent flights flown all other Wednesdays in March 2002. Non-stop, direct, single-connect and double-connect itineraries are generated using logic that simulates itinerary building rules used by computer reservation systems. Itinerary reports can differ from actual booked itineraries because: 1) an average week is used to represent all flights flown in a month, and 2) the connection logic does not accurately simulate itinerary building rules used to create bookings" (Garrow 2004). OAG data are publically available; however, the algorithms that are used to generate itineraries are typically proprietary (and thus researchers examining problems that use itinerary information typically need to develop their own itinerary-generation rules to replicate those found in practice).

Operations Data

There are many types of operational statistics and databases. For example, proprietary airline check-in data provide day of departure information from the passenger perspective, that is, it provides the ability to track passenger movements across flights and determine whether passengers show, no show, or successfully stand by for another flight. From a flight perspective, multiple proprietary and publically-available databases exist and contain information about flight departure delays and cancellations. For example, the U.S. DOT's Bureau of Transportation

Statistics (BTS) tracks on-time performance of domestic flights (Research and Innovative Technology Administration 2009) and provides high-level reasons for delays (weather, aircraft arriving late, airline delay, National Aviation System delay, security delay, etc.). Airlines typically maintain more detailed databases that track flights by their unique tail numbers and capture more detailed delay information (e.g., scheduled versus actual arrival and departure times at gate (or block times); scheduled versus actual taxi-in and taxi-out times; schedule versus actual time in flight, etc.) More detailed information on underlying causes associated with each delay component is also typically recorded (e.g., departure delay due to mechanical problem, late arriving crew, weather, etc.) When modeling air travel demand and air traveler behavior, it is useful to include operations information, identify flights and/or days of the year that have experienced unusually long delays and/or high flight cancellations (often due to weather storms, labor strikes, etc.) and exclude these data points from the analysis.

Summary of Main Concepts

This chapter presented one of the key motivations for writing this book: namely, the recent interest expressed by airlines and operations research analysts in modeling demand as the collection of individuals' choices using discrete choice models. Given that early applications and methodological developments associated with discrete choice models occurred predominately in the urban travel demand area, this chapter highlighted key distinctions between the operations research and urban travel demand areas. The most important concepts covered in this chapter include the following:

- In contrast to urban travel demand applications, aviation applications are characterized by relatively large volumes of revealed preference data that are used to produce demand forecasts that are a critical part of an airline's day-to-day operations. In this context, being able to measure both the accuracy and variability of forecasts is important.
- Data used to support an airline's day-to-day operations typically contain limited socio-economic and socio-demographic information.
- To date, the majority of aviation applications that have applied discrete choice models using revealed preference data fall into two main areas: 1) forecasts in which it is relatively easy to identify the set of alternatives an individual selects from; and 2) forecasts that are part of a larger decision-support system, but are "modularized" and easily replaceable. The applications form the basis of many of the examples presented in the text.
- Accurately representing competition among alternatives is important to both urban travel demand and aviation studies. Chapter 3 to 6 cover discrete choice methodological developments related to incorporating more flexible substitution patterns among alternatives. These developments, which

represent major milestones in the advancement of discrete choice theory, include the nested logit, generalized nested logit, and Network Generalized Extreme Value models.

- Many types of databases are available to support airline demand analysis and include booking, ticketing, schedule, and operations data. Typically, proprietary airline data contain more detailed information than data that are publically available. However, non-proprietary data that are commercially available or provided by government agencies are useful for understanding demand for air service across multiple carriers and markets.
- The U.S. is unique in that it is one of the few countries that collects a 10 percent ticket sample of passengers boarding domestic flights. This results in a valuable database that is used by both the practitioners as well as researchers.
- Due to the increased penetration of the Internet and subsequent increase in on-line and carrier-direct bookings, CRS booking databases that were previously valuable in determining market demands have become less reliable.

Chapter 2
Binary Logit and Multinomial Logit Models

Introduction

Discrete choice models, such as the binary logit and multinomial logit, are used to predict the probability a decision-maker will choose one alternative among a finite set of mutually exclusive and collectively exhaustive alternatives. A decision-maker can represent an individual, a group of individuals, a government, a corporation, etc. Unless otherwise indicated, the decision-making unit of analysis will be defined as an individual.

Discrete choice models relate to demand models in the sense that the total demand for a specific good (or alternative) is represented as the collection of choices made by individuals. For example, a binary logit model can be used to predict the probability that an airline passenger will no show (versus show) for a flight. The total demand expected to no show for a flight can be obtained by adding the no show probabilities for all passengers booked on the flight. This approach is distinct from statistical techniques traditionally used by airlines to model flight, itinerary, origin-destination, market, and other aggregate demand quantities. Probability and time-series methodologies that directly predict aggregate demand quantities based on archival data are commonly used in airline practice (e.g., demand for booking classes on a flight arrives according to a Poisson process, cancellations are binomially distributed, the no show rate for a flight is a weighted average of flight-level no show rates for the previous two months). In general, probability and time-series models are easier to implement than discrete choice models, but the former are limited because they do not capture or explain how individual airline passengers make decisions. Currently, there is a growing interest in applying discrete choice models in the airline industry. This interest is driven by the desire to more accurately represent *why* an individual makes a particular choice and *how* the individual makes trade-offs among the characteristics of the alternatives.

The interest in integrating discrete choice and other models grounded in behavioral theories with traditional revenue management, scheduling, and other applications is also being driven by several factors, including the increased market penetration of low cost carriers, wide-spread use of the Internet, elimination and/or substantial reduction in travel agency commissions, and introduction of simplified fare structures by network carriers. The presence of low cost carriers has reduced average market fares and increased the availability of low fares. Moreover, the Internet has reduced individuals' searching costs and made it easier for individuals to both find these fares and compare fares across multiple carriers without the

assistance of a travel agent. The elimination of commissions has removed the incentive of travel agencies to concentrate sales on those carriers offering the highest commissions. The introduction of simplied fare structures by network carriers was motivated by the need to offer products competitive with those sold by low cost carriers. Often, low cost carrier products do not require Saturday night stays and have few fare-based restrictions. However, these simplified fares have been less effective in segmenting price-sensitive leisure passengers willing to purchase weeks in advance of flight departure from time-sensitive business passengers willing to pay higher prices and needing to make changes to tickets close to flight departure. All of these factors have resulted in the need to better model how passengers make purchasing decisions, and to determine their willingness to pay for different service attributes. Moreover, unlike traditional models based solely on an airline's internal data, there is now a perceived need to incorporate existing and/or future market conditions of competitors when making pricing, revenue management, and other business decisions. Discrete choice models provide one framework for accomplishing these objectives.

This chapter presents fundamental concepts of choice theory and reviews two of the most commonly used discrete choice models: the binary logit and the multinomial logit models.

Fundamental Elements of Discrete Choice Theory

Following the framework of Domencich and McFadden (1975), it is common to characterize the choice process by four elements: a decision-maker, the alternatives available to the decision-maker, attributes of these alternatives, and a decision rule.

Decision-maker

A decision-maker can represent an individual (e.g., an airline passenger), a group of individuals (e.g., a family traveling for leisure), a corporation (e.g., a travel agency), a government agency, etc. Identifying the appropriate decision-making unit of analysis may be a complex task. For example, airlines often offer discounts to large corporate customers. As part of the discount negotiation process, airline sales representatives assess the ability of the corporation to shift high-yield trips from competitors to their airline. On one hand, the corporation's total demand is the result of thousands of independent travel decisions made by its employees. Employee characteristics (e.g., their membership and level in airlines' loyalty programs) and preferences (e.g., their preferences for aircraft equipment types, departure times, etc.) will influence the choice of an airline. In this sense, the decision-making unit of analysis is the individual employee. However, employees must also comply with their corporation's travel policies. In this sense, the corporation is also a decision-maker because it influences the choice of an airline

through establishing and enforcing travel policies. Thus, failure to consider the potential interactions between employee preferences and corporate travel policies may lead the sales representative to overestimate (in the case of weakly enforced travel policies) or underestimate (in the case of strongly enforced travel policies) the ability of the corporation to shift high-yield trips to a selected airline.

Alternatives

Each decision-maker is faced with a choice of selecting one alternative from a finite set of mutually exclusive and collectively exhaustive alternatives. Although alternatives may be discrete or continuous, the primary focus of this text is on describing methods applicable to selection of discrete alternatives. The finite set of all alternatives is defined as the *universal choice set*, C. However, individual n may select from only a subset of these alternatives, defined as the *choice set*, C_n. In an itinerary choice application, the universal choice set could be defined to include all reasonable itineraries in U.S. markets that depart from cities in the eastern time zone and serve cities in the western time zone, whereas the choice set for an individual traveling from Boston to Portland would contain only the subset of itineraries between these two city pairs. In practice, the universal choice set is often defined to contain only reasonable alternatives. In itinerary choice applications, distance-based circuitry logic can be used to eliminate unreasonable itineraries and minimum and maximum connection times can be used to ensure that unrealistic connections are not allowed.

There are several subtle concepts related to the construction of the universal choice set. First, the assumptions that alternatives are mutually exclusive and collectively exhaustive are generally not restrictive. For example, assume there are two shops in an airport concourse, a dining establishment and a newsstand, and an airport manager is interested in knowing the probability an airline passenger will make a purchase at one or both of these stores. The choice set cannot be defined using simply two alternatives, as they are not mutually exclusive, *i.e.*, the passenger can choose to shop in both stores. Mutual exclusivity can be obtained using three alternatives: “purchase only at dining establishment,” “purchase only at newsstand,” and “purchase both at dining establishment and newsstand.” To make the choice set exhaustive, a fourth alternative representing customers who “do not purchase” can be included.

Also, the way in which the universal choice set is defined can lead to different interpretations. Consider a situation in which the analyst wants to predict the probability an individual will select one of five itineraries serving a market. The universal choice set is defined to contain these five itineraries, $C_1 \in \{I_1, I_2, I_3, I_4, I_5\}$, and a discrete choice model calibrated using *actual booking data* is used to predict the probability that one of these alternatives is selected. Compare this to a situation in which the analyst has augmented the universal choice set to include a no purchase option, $C_2 \in \{I_1, I_2, I_3, I_4, I_5, NP\}$, and calibrates the choice model using *booking requests* that are assumed to be independent. The first model will predict the

probability an individual will select a particular itinerary given that the individual has decided to book an itinerary. The second will predict both the probability that an individual requesting itinerary information will purchase an itinerary, $1 - \Pr(NP)$, and, if so, which one will be purchased. The probability that itinerary one will be chosen out of all booking requests is given as $\Pr(I_1)$ and the probability that itinerary one will be chosen out of all bookings is $\Pr(I_1)/\{1 - \Pr(NP)\}$. This example demonstrates how different interpretations can arise from seemingly subtle changes in the universal choice set. It also illustrates how data availability can influence the construction of the universal choice set.

Attributes of the Alternatives

The third element in the choice process defined by Domencich and McFadden (1975) refers to attributes of the alternatives. Attributes are characteristics of the alternative that individuals consider during the choice process. Attributes can represent both deterministic and stochastic quantities. Scheduled flight time is deterministic whereas the variance associated with on-time performance is stochastic. In itinerary choice applications, attributes include schedule quality (non-stop, direct, single connection, double connection), connection time, departure and/or arrival times, aircraft type, airline, average fare, etc. In practice, the attributes used in scheduling, revenue management, pricing, and other applications that support day-to-day airline operations are derived from *revealed preference* data. Revealed preference data are based on the actual, observed behavior of passengers. By definition, revealed preference data reflect passenger behavior under existing or historical market conditions. Internal airline data rarely contain gender, age, income, marital status or other socio-demographic information. Passenger information is generally limited to that collected to support operations. This includes information about the passenger's membership and status in the airline's loyalty program as well as any special service requests (e.g., wheelchair assistance, infant-in-arms, unaccompanied minor, special meal request).

When developing models of airline passenger behavior, it is desirable to identify which attributes individuals consider during the choice process and how passengers value these attributes according to trip purpose and market. Intuitively, leisure passengers will tend to be more price-sensitive and less time-sensitive than business passengers. Given that trip purpose is not known, heterogeneity in customers' willingness to pay is achieved by using proxy variables to represent trip purpose. These include the number of days in advance of flight departure a booking is made, departure day of week and length of stay, presence of a Saturday night stay, flight departure and/or arrival times, number of passengers traveling together on the same reservation, etc. Compared to leisure passengers, business travelers tend to book close to flight departure, travel alone during the most popular times of day, depart early in the work week and stay for shorter periods, and avoid staying over a Saturday night. However, day of week, time of day, and other preferences will vary by market. A business traveler wanting to arrive for

a Monday meeting in Tokyo may prefer a Friday or Saturday departure from the U.S. to recover from jet lag, whereas a business traveler departing from Boston to Chicago for a Monday meeting may prefer to depart early Monday morning to spend more time at home with family.

When modeling air traveler behavior, it is important to account for passenger preferences across markets. One common practice is to group "similar" markets into a common dataset and estimate separate models for each dataset. Similarity is often defined according to the business organization of the airline. For example, a domestic U.S. carrier may have several groups of pricing analysts, each responsible for a group of markets (Atlantic, Latin, Pacific, domestic hub market(s), leisure Hawaii and Florida markets, etc.). Alternatively, similarity may be defined using statistical approaches like clustering algorithms.

Although revealed preference data are used in the majority of airline applications, there are situations in which inferences from revealed preference data are of limited value. The exploration of the effects of new and non-existent service attributes, such as new cabin configurations and new aircraft speeds and ranges, is a critical component of Boeing's passenger modeling. Moreover, the inclusion of passenger social, demographic and economic variables in the model formulations are vital to understanding what motivates and segments passenger behavior across different regions of the world. These data are rarely, if ever, available in revealed preference contexts. Consequently, Boeing's and other company's marketing departments invest millions in *stated preference* surveys and mock-up cabins when designing a new aircraft (Garrow, Jones and Parker 2007).

Model enhancements are often driven by the need to include additional attributes to support or evaluate new business processes. For example, prior to the use of code-shares, there was no need to distinguish between the marketing carrier who sold a ticket and the carrier who operated the flight, as these were the same carrier. In order to predict incremental revenue associated with an airline entering into different code-share agreements, it was necessary to model how itineraries marketed as code-shares differed from those marketed and flown by the operating carrier. When prioritizing model enhancements, a balance needs to be obtained between making models complex enough to capture factors essential for accurately supporting and evaluating different "what-if scenarios" while making these models simple enough to be understood by users and flexible enough to incorporate new attributes that were not envisioned when the model was first developed.

Decision Rule

The final element of the choice process is the decision rule. Numerous decision rules can be used to model rational behavior. Following the definition of Ben-Akiva and Lerman (1985), rational behavior refers to an individual who has consistent and transitive preferences. Consistent preferences refer to the fact that an individual will consistently choose the same alternative when presented with two identical choice situations. Transitive preferences capture the fact that if alternative A is

preferred to alternative B and alternative B is preferred to alternative C then alternative A is preferred to alternative C.

Ben-Akiva and Lerman (1985) categorize decision rules into four categories: dominance, satisfaction, lexicographic, and utility. Figure 2.1 portrays time and cost attributes associated with five alternatives. Note the definition of the axis, which places the most attractive alternatives (those with least time and cost) in the upper right. The dominance rule eliminates alternatives that are clearly inferior (i.e., that have both higher time and cost than another alternative). Formally, alternative i dominates alternative j, if and only if $x_{ik} \geq x_{jk}\ \forall k$, where k represents the vector of attributes. When using the dominance rule, alternatives B and D are eliminated. The time and cost associated with alternative B are both larger than those associated with alternative C. Similarly, the time and cost associated with alternative D are both larger than those associated with alternative E. Alternatives A, C, and E remain in the non-dominated set of solutions. This highlights two of the major limitations of using a dominance decision rule for choice theory. Specifically, application of the dominance rule may not lead to a single, unique choice and it does not capture how individuals make trade-offs among attributes.

Satisfaction and lexicographic decision rules are also limited in the sense that they do not capture how individuals make trade-offs among attributes and can result in non-unique choices. According to the satisfaction decision rule, all alternatives that satisfy a minimum requirement (S_k) for all attributes are retained for consideration. Formally, alternative i is retained for consideration *iff* $x_{ik} \geq S_k\ \forall k$. Figure 2.2 illustrates the application of the satisfaction decision rule. Alternatives

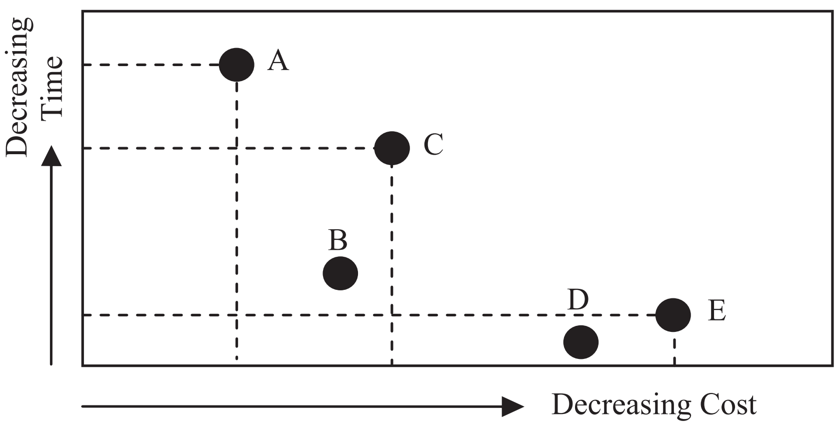


Figure 2.1 Dominance rule

Source: Adapted from Koppelman 2004: Figure 1.1 (reproduced with permission of author).

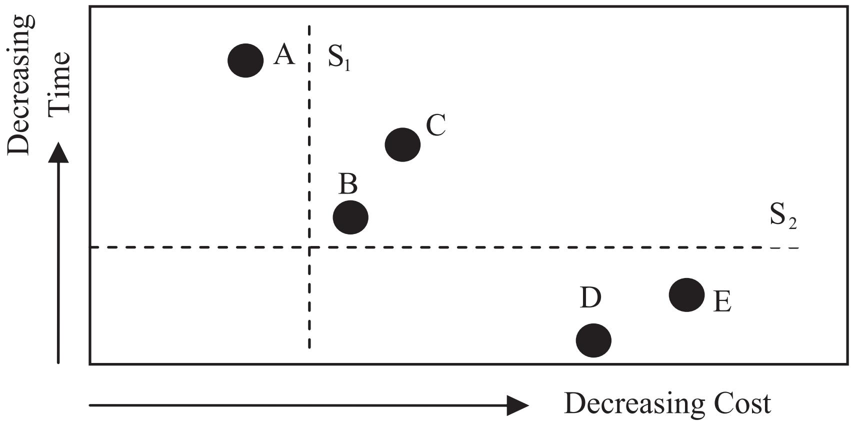


Figure 2.2 Satisfaction rule

Source: Adapted from Koppelman 2004: Figure 1.2 (reproduced with permission of author).

B and C will be retained for choice consideration as they are the only alternatives that have costs less than S_1 and travel times less than S_2. The satisfaction rule can be used to simplify the choice scenario by screening alternatives to include in the choice set.

According to the lexicographic decision rule, attributes are first ordered by importance and the alternative(s) with the highest value for the most important attribute is selected. If the choice is not unique, the process is repeated for the second most important attribute. The process is repeated until only one alternative remains. Formally, select all i alternatives such that $x_{i1} \geq x_{j1} \; \forall j \in C_n$. If the remaining choice set, C_n^i, is not unique, select all l alternatives, such that $x_{l2} \geq x_{i2} \; \forall l \in C_n^i$ and repeat the process until only one alternative remains. Consider the five alternatives shown in Table 2.1. Assuming time is the most important attribute, alternatives A, B, and C would be considered. Assuming cost is the second most important attribute, alternatives A and B would be considered. Note that alternative D and E cannot be chosen, although they have the lower costs than alternatives A and B, because they were eliminated in the first round. Finally, assuming seat location is the third most important attribute and the passenger prefers an aisle, alternative B would be the one ultimately selected. The example highlights one of the main problems with the lexicographic rule, i.e., the ordering of the importance of attributes can be subjective and does not enable the individual to make trade-offs among the attributes.

The final category of decision rules is based on the concept of utility. Utility is a scalar index of value that is a function of attributes and/or individual characteristics. In contrast to the other decision rules, utility represents the "value" an individual places on different attributes and captures how individuals

Table 2.1 Lexicographic rule

	Time	Cost	Seat
Alt A	30	$200	Window
Alt B	30	$200	Aisle
Alt C	30	$250	Middle
Alt D	60	$100	Middle
Alt E	75	$150	Aisle

make trade-offs among different attributes. Individuals are assumed to select the alternative that has the maximum utility. Alternative i is chosen if the utility individual n obtains from alternative i, U_{ni}, is greater than the utility for all other alternatives. Formally, alternative i is chosen $iff\ U_{ni} > U_{nj}\ \forall j \neq i$. The utility for alternative i and individual n, U_{ni}, has an observed component, V_{ni}, and an unobserved component, commonly referred to as an "error term," ε_{ni}, but is more precisely referred to as the "stochastic term." Formally, $U_{ni} = V_{ni} + \varepsilon_{ni}$, where $V_{ni} = \boldsymbol{\beta}' x_{ni}$. The observed component is often called the systematic or representative component of utility. The observed component is typically assumed to be a linear-in-parameters function of attributes that vary across individuals and alternatives (e.g., price, flight duration, gender). Note the assumption that $\boldsymbol{\beta}$ is linear-in-parameters *does not* imply that attributes like price must have linear relationships, i.e., x_{ni} can take on different functional forms, such as price, log (price), price^2; a linear-in parameters assumption means that just the coefficient, $\boldsymbol{\beta}$, associated with x_{ni} must be linear.

The error component is a random term that represents the unobserved and/or unknown (to the analyst) portion of the utility function. The distribution of random terms may be influenced by several factors, including measurement errors, omitting attributes from the utility function that are important to the choice process but that cannot be measured and/or are not known, incorrectly specifying the functional form of attributes that are included in the model (e.g., using a linear relationship when the "true" relationship is non-linear), etc. There is an implicit relationship between the attributes included in the model and the distribution of error terms. That is, by including different attributes and/or by changing how attributes are included in the model, the distribution of error terms may change. Conceptually, this is similar in spirit to the situation where an analyst specifies a linear regression model and then examines the distribution of residual errors using visual plots and/or statistical tests to ensure that homoscedasticity and other assumptions embedded in the linear regression model are maintained. However, because choice models predict the probabilities associated with multiple, discrete outcomes, the ability to visually assess the appropriateness of error distribution assumptions is limited. Consequently, discrete choice

modeling relies on statistical tests to identify violations in assumptions related to error distributions (e.g., see Train 2003: 53-4 for an extensive discussion and review of these tests). In addition, it is common to estimate different models (derived from different assumptions on the error terms) as part of the modeling process and assess which model fits the data the best.

Derivation of Choice Probabilities and Motivation for Different Choice Models

One of the first known applications of a discrete choice model to transportation occurred in the early 1970's when Daniel McFadden used a multinomial logit formulation to model mode choice in the San Francisco Bay Area. Since the 1970's, dozens of discrete choice models[1] have been estimated and applied in transportation, marketing, economics, social science, and other areas. This section presents the general methodology used to derive choice probabilities for these models and describes how limitations of early discrete choice models motivated the development of more flexible discrete choice models.

The derivation of choice probabilities for discrete choice models uses the fact that individuals are assumed to select the alternative that has the maximum utility. Specifically, the utility associated with alternative i for individual n is given as $U_{ni} = V_{ni} + \varepsilon_{ni}$ and the probability the individual selects the alternative i from all J alternatives in the choice set C_n is given as:

$$\begin{aligned}
P_{ni} &= P\left(U_{ni} \geq U_{nj} \forall\, j \neq i\right) \\
&= P\left(V_{ni} + \varepsilon_{ni} \geq V_{nj} + \varepsilon_{nj} \forall\, j \neq i\right) \\
&= P\left(\varepsilon_{nj} - \varepsilon_{ni} \leq V_{ni} - V_{nj} \forall\, j \neq i\right) \\
&= P\left(\varepsilon_{nj} \leq V_{ni} - V_{nj} + \varepsilon_{ni} \forall\, j \neq i\right) \\
&= \int_{\varepsilon_i=-\infty}^{+\infty} \int_{\varepsilon_j=-\infty\ \forall j\neq i}^{V_i - V_j + \varepsilon_i} f\left(\varepsilon\right) d\varepsilon_J, \ldots, d\varepsilon_{i+1}, d\varepsilon_i
\end{aligned}$$

This derivation is general in the sense that no assumptions have been made on the distribution of error terms; these assumptions are required in order to derive choice probabilities for specific models. However, the general derivation illustrates that the probability an alternative is selected is a function of both the observed and unobserved components of utility. This means that even though the

1 Here, the term "model" is used to refer to the formulas used to compute choice probabilities. Examples of different "models" include the binary logit, multinomial logit, nested logit, and mixed logit.

observed utility for alternative i is greater than the *observed* utility for alternative j, alternative j may still be chosen. This will occur when the *unobserved* utility for alternative j is "sufficiently larger" than the *unobserved* utility for alternative i, $P_{ni} = P\left(\varepsilon_{nj} \leq V_{ni} - V_{nj} + \varepsilon_{ni} \forall j \neq i \right)$. The probability that ε_{nj} is less than $(V_{ni} - V_{nj} + \varepsilon_{ni})$ is obtained from the cumulative distribution function (cdf), i.e., by integrating over the joint probability distribution function of error terms, $f(\boldsymbol{\varepsilon})$. Because the cdf is continuous, the case in which the utility of the two alternatives is identical, $U_{ni} = U_{nj}$, is irrelevant to the derivation of choice probabilities.

Specific choice probabilities for different discrete choice models are obtained by imposing different assumptions on the distribution of these error terms. The assumption that unobserved error components are independently and identically distributed (iid) and follow a Gumbel distribution with mode zero and scale one, $\boldsymbol{\varepsilon} \sim$ iid G (0,1), results in the binary logit (in the case of two alternatives) or the multinomial logit model (in the case of more than two alternatives) (McFadden 1974). The assumption that the error terms are iid G(0,1) is advantageous in the sense that the choice probability takes on a closed-form expression that is computationally simple. However, the same assumption imposes several restrictions on the binary logit and multinomial logit (MNL) models. First, the assumption that error terms are iid across alternatives leads to the independence of irrelevant alternatives (IIA), a property which states that the ratio of choice probabilities P_{ni} / P_{nj} for $i, j \in C_n$ is independent of the attributes of any other alternative. In terms of substitution patterns, this means a change or improvement in the utility of one alternative will draw share proportionately from all other alternatives. In many applications, this may not be a realistic assumption. For example, in itinerary choice model applications, one may expect the 10 AM departure to compete more with flights departing close to 10 AM. Second, the assumption that error terms are iid across observations restricts correlation among observations. This is not a realistic assumption when using data that contain multiple responses from the same individual (e.g., when using panel data or multiple-response survey data or online search data that span multiple visits by the same individual). Third, the assumption that error terms are identically distributed across alternatives and individuals implies equal variance, or homoscedasticity. This may not be a realistic assumption when the variance of the unobserved portion of utility is expected to vary as a function of another variable. For example, in mode choice models the variance associated with travel time is expected to increase as a function of distance.

A fourth limitation of MNL models is that they cannot incorporate *unobserved* random taste variation. Observed taste variation can be directly incorporated into model specifications by including individual socio-economic characteristics as alternative-specific variables or by interacting these variables with generic variables describing the attributes of each alternative. (A classic example is to define sensitivity of cost as a decreasing function of an individual's income.) The MNL model (as well as all models with fixed coefficients) assumes that the

$\boldsymbol{\beta}$ coefficients in the utility function associated with observable characteristics of alternatives and individuals are fixed over the population. Models that incorporate unobserved random taste such as the mixed logit model allow the $\boldsymbol{\beta}$ coefficients to vary over the population. As described by Jain, Vilcassim, and Chintagunta (1994) and Bhat and Castelar (2002), unobserved random taste variation can be classified as preference heterogeneity or response heterogeneity. Preference heterogeneity allows for differences in individuals' preferences for a choice alternative (preference homogeneity implies that individuals with the same observed characteristics have identical choice preferences). Response heterogeneity allows for differences in individual's sensitivity or "response" to characteristics of the choice alternatives. In practice, preference heterogeneity is modeled by allowing the alternative specific constants (or intercept terms) to vary over the population whereas response heterogeneity is modeled by allowing parameters associated with individual or alternative specific characteristics to vary over the population. As a side note, the mixed logit chapter shows how imposing distributional assumptions on the $\boldsymbol{\beta}$ coefficients is equivalent to imposing distributional assumptions on error terms; thus, the earlier statement that different choice models are derived via distributional distribution assumptions on error components is accurate. From a practical interpretation perspective, it is more natural to frame random taste variation in the context of the $\boldsymbol{\beta}$ coefficients.

Although the assumption that error terms are iid G(0,1) leads to the elegant, yet restrictive MNL model, the assumption that the error terms follow a multivariate normal distribution with mean zero and covariance matrix $\sum_{\varepsilon}$, $\varepsilon \sim MVN\,(0,\sum_{\varepsilon})$, results in the multinomial probit (MNP) model (Daganzo 1979). Unlike the MNL, the probit model allows flexible substitution patterns, correlation among unobserved factors, heteroscedasticity, and random taste variation. However, the choice probabilities can no longer be expressed analytically in closed-form and must be numerically evaluated.

Conceptually, MNL and MNP models can be loosely thought of as the endpoints of a spectrum of discrete choice models. On one end is the MNL, a restrictive model that has a closed-form probability expression that is computationally simple. On the other end is the MNP, a flexible model that has a probability expression that must be numerically evaluated. Over the last 35 years, advancements in discrete choice models have generally focused on either relaxing the substitution restriction of the MNL while maintaining a closed-form expression for the choice probabilities or reducing the computational requirements of open-form models and further expanding the spectrum of open-form models to include more general formulations. This text focuses on those closed-form and open-form discrete choice models that are most applicable to the study of air travel demand. For additional references, see Koppelman and Sethi (2000) and Koppelman (2008) for reviews of closed-form advancements and Bhat (2000a) and Bhat, Eluru, and Copperman (2008) for reviews of open-form advancements.

Properties of the Gumbel Distribution

The assumption that error terms are Gumbel (or Extreme Value Type I) distributed is common to many choice models, including the binary logit, multinomial logit, nested logit, cross-nested logit, and generalized nested logit. This section presents some of the most important properties of the Gumbel distribution. These properties are used to derive different discrete choice models. Knowing how to use these properties to derive different choice models is not essential to learning how to interpret and apply discrete choice models. However, these same properties influence the interpretation of choice probabilities in many subtle, yet important ways. In addition, a thorough understanding of these concepts is often required to apply choice models in a research context. Thus, there is tremendous benefit in mastering the subtle concepts related to the properties of the Gumbel distribution and understanding how these properties are meaningfully connected to the interpretation of choice probabilities. For these reasons, the properties of the Gumbel distribution are emphasized from the beginning of the text, and the relationships between these properties and the interpretation of choice model probabilities are explicitly detailed. For a more comprehensive overview of the properties of the Gumbel distribution beyond those presented here, see Johnson, Kotz, and Balakrishnan (1995).

Cumulative and Probability Distribution Functions

The cumulative distribution function (cdf) and probability distribution function (pdf) of the Gumbel distribution are given as:

$$F(\varepsilon) = \exp\left\{-\exp\left[-\gamma(\varepsilon - \eta)\right]\right\}, \gamma > 0$$

$$f(\varepsilon) = \gamma \times \exp\left[-\gamma(\varepsilon - \eta)\right] \times \exp\left\{-\exp\left[-\gamma(\varepsilon - \eta)\right]\right\}$$

where η is the mode and γ is the scale. Unlike the normal distribution, the Gumbel is not symmetric and its distribution is skewed to the right, which results in its mean being larger than its mode. The mean and variance of the Gumbel distribution are obtained from the following relationships:

$$\text{mean} = \eta + \frac{\text{Euler constant}}{\gamma} \approx \eta + \frac{0.577}{\gamma}$$

$$\text{variance} = \frac{\pi^2}{6\gamma^2}$$

Note that unless otherwise stated, this text defines the scale of the Gumbel distribution with respect to the "inverse variance." That is, given the scale, $\gamma > 0$, the variance is defined as $\pi^2/(6\gamma^2)$. Some researchers define the relationship between the scale and variance as $\pi^2\gamma^2/6$. The choice of whether to define variance

using the "scale" or "inverse scale" relationship is somewhat arbitrary, although in some derivations, one definition may be easier to work with than the other. However, because different definitions exist (and can easily be confused), it is important to explicitly note how the scale parameter relates to the definition of variance.

Although the Gumbel is not symmetric, it is very similar to the normal distribution. The similarity can be seen in Figures 2.3 and 2.4. The mean and variance of the Gumbel and normal distributions in the figures are identical. The mean of the Gumbel distribution is $2 + 0.5773/3 = 2.19$, the variance is $\pi^2 / (6 \times 3^2) = 0.183$, and standard deviation is $0.183^{0.5} = 0.43$.

Scale and Translation of the Gumbel Distribution

Assume $\varepsilon \sim G(\eta, \gamma)$ and Z and ω are constants > 0. The sum of $(\varepsilon + Z)$ also follows a Gumbel distribution with the same scale, but its mode will be shifted (or "translated") by Z units. Formally, $(\varepsilon + Z) \sim G(\eta + Z, \gamma)$. Multiplying ε by a constant will also result in a Gumbel distribution, albeit with both a different mode and scale: $\omega\varepsilon \sim G(\omega\eta, \gamma / \omega)$. These properties are illustrated in Figure 2.5. Just as the unit normal distribution can be used as a reference for more general

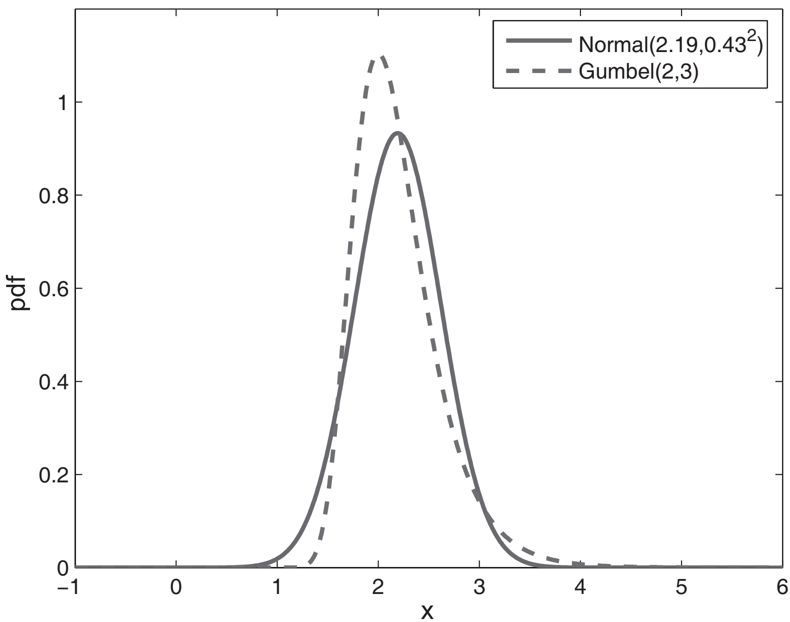


Figure 2.3 PDF for Gumbel and normal (same mean and variance)

Source: Adapted from Koppelman 2004: Figure 2.1 (reproduced with permission of author).

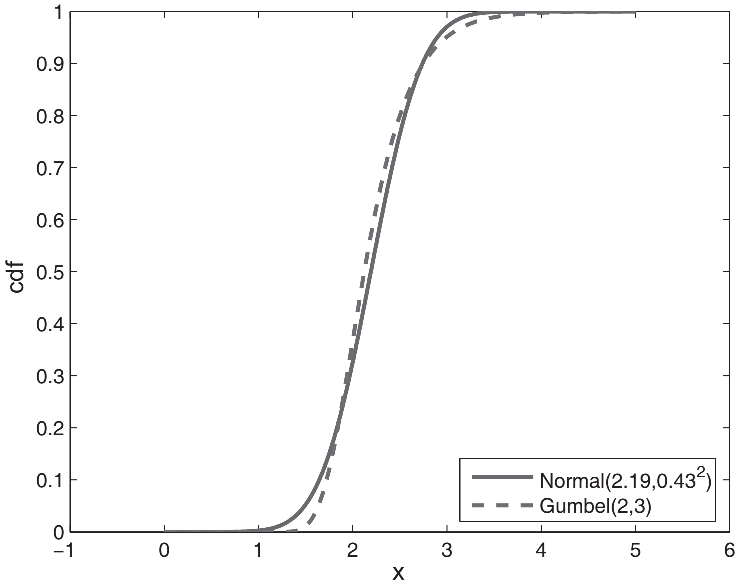


Figure 2.4 CDF for Gumbel and normal (same mean and variance)

Source: Adapted from Koppelman 2004: Figure 2.2 (reproduced with permission of author).

normal distributions, so can the unit Gumbel. That is, any Gumbel distribution can be formed from a unit Gumbel distribution by applying scale and translation adjustments.

Difference of Two Independent Gumbel Random Variables with the Same Scale

Assume ε_1 and ε_2 are independently distributed Gumbel such that they have the same scale, but different modes. Formally, $\varepsilon_1 \sim G(\eta_1,\gamma)$ and $\varepsilon_2 \sim G(\eta_2,\gamma)$. Then, $\varepsilon^* = (\varepsilon_2 - \varepsilon_1)$ is logistically distributed with cdf and pdf:

$$F\left(\varepsilon^*\right) = \frac{1}{1 + \exp\left[\gamma\left(\eta_2 - \eta_1 - \varepsilon^*\right)\right]}$$

$$f\left(\varepsilon^*\right) = \frac{\gamma \exp\left[\gamma\left(\eta_2 - \eta_1 - \varepsilon^*\right)\right]}{\left(1 + \exp\left[\gamma\left(\eta_2 - \eta_1 - \varepsilon^*\right)\right]\right)^2},\ \gamma > 0$$

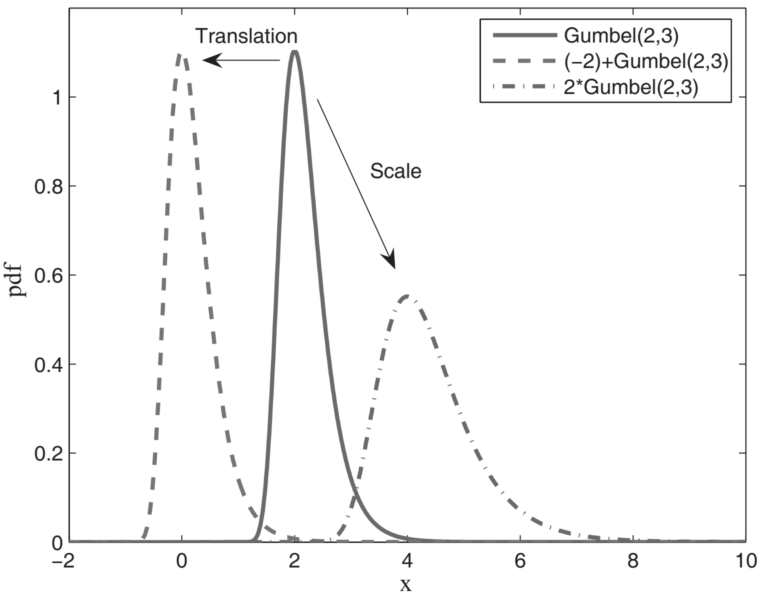


Figure 2.5 Scale and translation of Gumbel

where η is the mode and γ is the scale. The logistic distribution is symmetric and its mean, mode, and variance are given as:

$$\text{mean} = \text{mode} = (\eta_2 - \eta_1)$$

$$\text{variance} = \frac{\pi^2}{3\gamma^2}$$

An example is provided in Figure 2.6. The first two panels depict the histograms of two Gumbel random variables G1 and G2, each with 1,000,000 observations. The first Gumbel random variable is distributed with mode three and scale one and the second Gumbel random variable is distributed with mode five and scale one. Consistent with the proof given in Gumbel (1958), the result of the difference (G2-G1) follows a logistic distribution with theoretical parameters of two for the location and one for the scale.

The cdf of the logistic distribution is also similar to the cdf of the Gumbel distribution, as shown in Figure 2.7. The mean and variance of the logistic and Gumbel distributions in the figure are identical.

Difference of Two Independent Gumbel Random Variables with Different Scales

While the difference of two independent Gumbel random variables with the same scale (and variance) follows a logistic distribution, the same cannot be

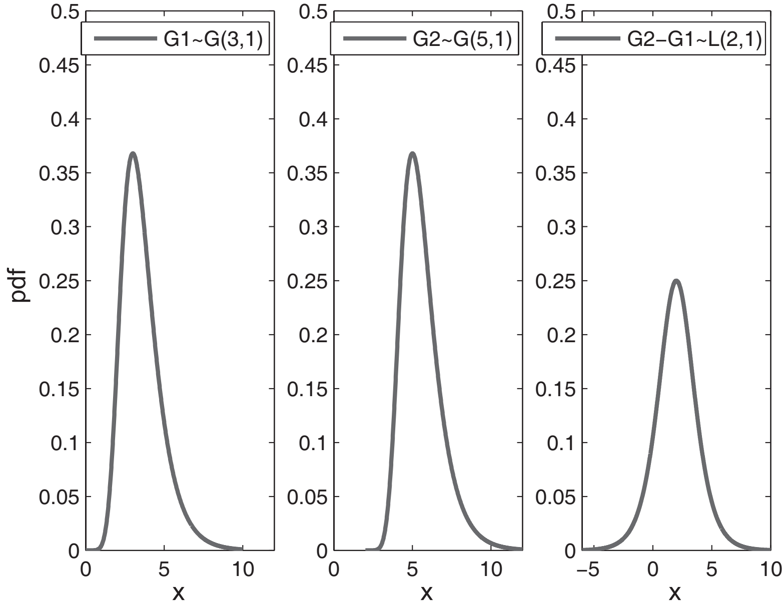


Figure 2.6 Difference of two Gumbel distributions with the same scale parameter

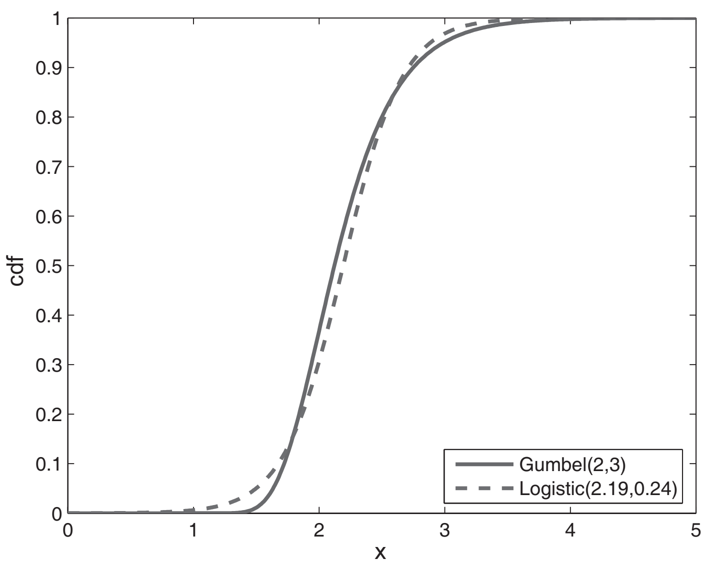


Figure 2.7 CDF for Gumbel and logistic (same mean and variance)

Source: Adapted from Koppelman 2004: Figure 2.4 (reproduced with permission of author).

said about the difference of two independent Gumbel random variables that have different scales. In this case, if the variance of one of the random variable is "large" compared to variance of the second random variable, the difference will asymptotically converge to a Gumbel distribution. Conceptually, this is because the random variable with the smaller variance behaves as a constant. That is, given $\varepsilon_1 \sim G(\eta_1,\gamma_1)$ and $\varepsilon_2 \sim G(\eta_2,\gamma_2)$ with $\gamma_2 \gg \gamma_1$ (which implies the variance of $\varepsilon_1 \gg \varepsilon_2$), $\varepsilon^* = (\varepsilon_2 - \varepsilon_1) \sim G(\eta_2 - \eta_1, \gamma_1)$. The problem arises in precisely defining what constitutes a "large difference in scale parameters" and characterizing the distribution that represents the case when the scales are slightly different. Early work by E. J. Gumbel (1935, 1944, 1958) discusses the problem, but it was not until 1997 that Cardell derived these pdf and cdf functions. Further, although Cardell shows that, under certain conditions, closed-form results can be obtained, the use of these pdf and cdf functions are generally limited due to the inability to efficiently operationalize them. Chapter 3, which covers the nested logit (NL) model, will revisit this issue in the context of how to generate synthetic NL datasets.

Figure 2.8 shows the histograms of two Gumbel random variables, each with 1,000,000 observations, which have the same location parameter but different scale parameters. Note the *y*-axis of the second panel ranges from 0 to 4 and the *y*-axis of the first and third panel ranges from 0 to 0.5. Because the ratio of the scale parameters is small, it is expected that (G_2-G_1) will follow a Gumbel distribution with the scale parameter of the distribution with the maximum variance, or G_1.

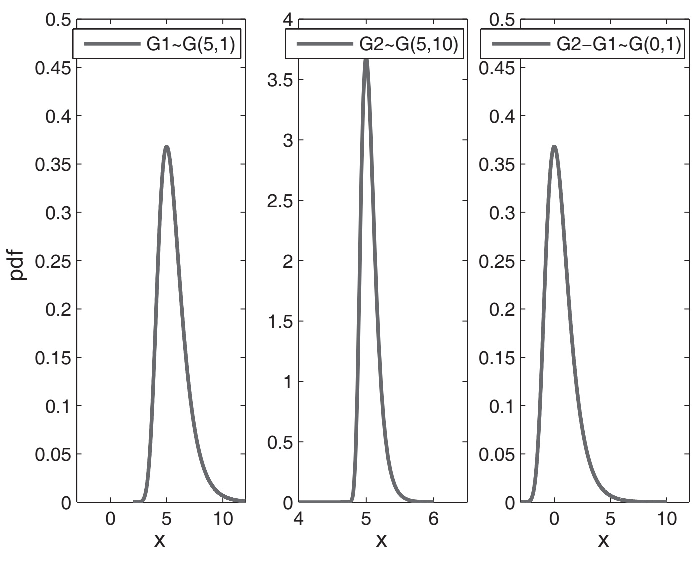


Figure 2.8 Difference of two Gumbel distributions with different scale parameters

Maximization over Independent Gumbel Random Variables

Assume ε_1 and ε_2 are independently distributed Gumbel such that they have the same scale, but different modes: $\varepsilon_1 \sim G(\eta_1,\gamma)$ and $\varepsilon_2 \sim G(\eta_2,\gamma)$. Then:

$$\max(\varepsilon_1,\varepsilon_2) \sim G\left(\frac{1}{\gamma}\ln\left(\exp(\gamma\eta_1)+\exp(\gamma\eta_2)\right),\gamma\right)$$

The results can be extended to maximize over J independently distributed Gumbel variables that have the same scale:

$$\max_j(\varepsilon_j) \sim G\left(\frac{1}{\gamma}\ln\sum_{j=1}^{J}\exp(\gamma\eta_j),\gamma\right)$$

An example is shown in Figure 2.9 for $\varepsilon_1 \sim G(3,1)$ and $\varepsilon_2 \sim G(4,1)$. The maximum of these two distributions is distributed $G(1\text{n}\{\exp(3)+\exp(4)\},1) = G(4.31,1)$.

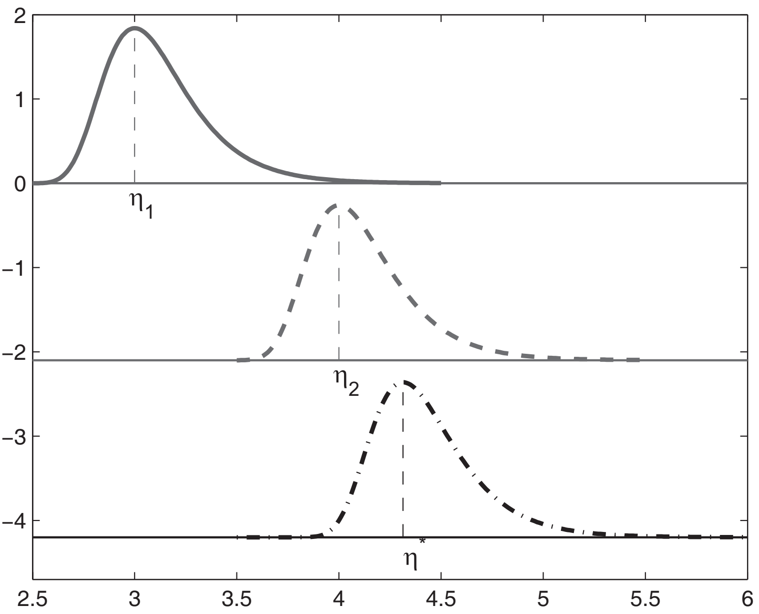


Figure 2.9 Distribution of the maximum of two Gumbel distributions (same scale)

Source: Adapted from Koppelman 2004: Figure 2.5 (reproduced with permission of author).

Why Do We Care About the Properties of the Gumbel Distribution?

The assumption that error terms follow a Gumbel distribution is common to many discrete choice models, including those that are most often used in practice. Although the properties discussed above may appear straightforward, they influence choice models in subtle ways. The next section describes how the properties of the Gumbel distribution influence the interpretation of choice probabilities.

Binomial Logit

Choice Probabilities

The binary logit model is used to describe how an individual chooses between two discrete alternatives. Consistent with maximum utility theory, the systematic or observable utility associated with alternative i for individual n is given as $U_{ni} = V_{ni} + \varepsilon_{ni}$ and the individual is assumed to choose the alternative with the maximum utility. Binary logit probabilities are derived from assumptions on error terms. Specifically, error terms are assumed to be iid Gumbel. As discussed in the previous section, under the assumption that ε_1 and ε_2 are iid G(0,1), $\varepsilon_2 - \varepsilon_1$ is logistically distributed. The binary logit probabilities take a form that is similar to the cdf of the logistic distribution:

$$P_{ni} = P\left(U_{ni} \geq U_{nj}\right)$$

$$P_{ni} = P\left(\varepsilon_{nj} - \varepsilon_{ni} \leq V_{ni} - V_{nj}\right)$$

$$P_{ni} = \frac{1}{1 + \exp\left[-\left(V_{ni} - V_{nj}\right)\right]}$$

A second common probability expression for the binary logit is obtained by multiplying the numerator and denominator by exp (V_{ni}), or

$$P_{ni} = \frac{\exp\left(V_{ni}\right)}{\exp\left(V_{ni}\right) + \exp\left(V_{nj}\right)}$$

There is an underlying sigmoid or S-shape relationship between observed utility and choice probabilities, as shown in Figure 2.10. The S-shape implies that an improvement in the utility associated with alternative i will have the largest impact on choice probabilities when there is an equal probability that alternatives i and j will be selected. That is, when the utilities (or values) of two alternatives are similar, improving one of the alternatives will have a larger impact on attracting customers from competitors. The relationship between service improvements and existing market position is a subtle point, yet one that is

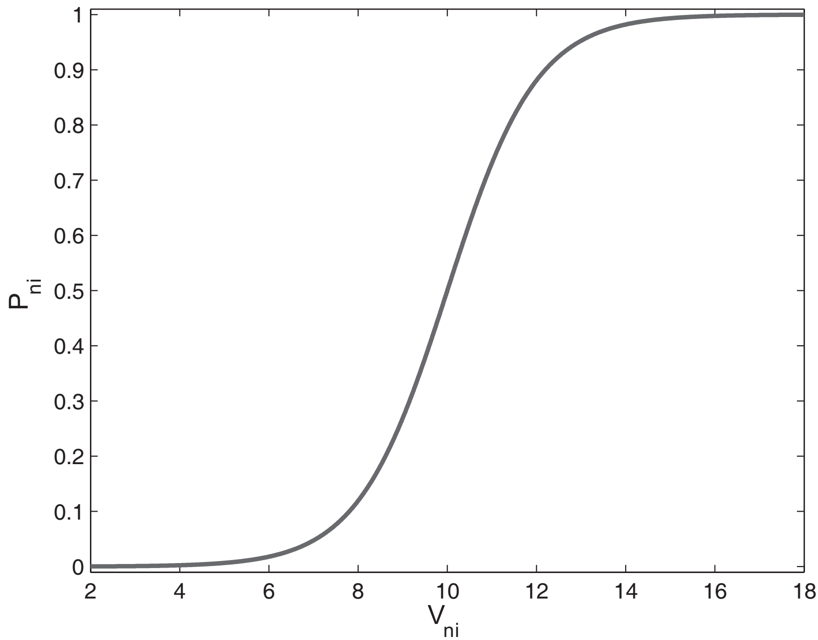


Figure 2.10 Relationship between observed utility and logit probability

important to consider when making large infrastructure or service improvements. The next sections describe two other properties of choice models that influence the interpretation of choice probabilities. Specifically, these sections describe why only differences in utility are uniquely identified and explain how choice probabilities and $\boldsymbol{\beta}$ parameter estimates are affected by the amount of variance associated with the unobserved portion of utility. The discussion of the binary logit model concludes with a discussion of the similarity between binary logit and logistic regression models. An emphasis is placed on showing how one of the common methods used in logistic regression models to interpret parameter estimates (specifically, odds ratios and enhanced odds ratios) can be applied to binary and multinomial logit models. Given that the derivation of choice probabilities for the binary and multinomial logit model provides limited value to the understanding of how to interpret choice models (but is helpful for those who plan to do research in this area), it is included as an appendix at the end of this chapter.

Only Differences in Utilities are Identified

The first binary logit formula illustrates that only differences in utilities are uniquely identified. Intuitively, "lack of identification" in this context means that

adding (or subtracting) a constant value to the utility of each alternative will not change the probability an alternative is selected. This fact must be taken into account when specifying both the systematic portion of utility and the unobserved portion of utility.

To illustrate the fact that only differences in utility are uniquely identified, consider a situation in which an individual chooses between two itineraries. The utility associated with itinerary *i* is a function of the number of stops and price (or fare) expressed in hundreds of dollars:

$$V_i = -0.4\left(price_i\right) - 0.5\left(stops_i\right)$$

Table 2.2 shows the utility calculations for two choice scenarios. In the first scenario, the individual must choose between a non-stop itinerary offered at \$700 and a one-stop itinerary offered at \$600. In the second scenario, the individual must choose between a one-stop itinerary offered at \$600 and a two-stop itinerary offered at \$500. The second scenario differs from the first in that the price of each itinerary is lowered by the same amount (\$100) and the number of stops of each itinerary is raised by the same amount (one stop). The difference $(V_1 - V_2) = 0.1$ is identical for both choice scenarios. The corresponding probabilities are also identical. Using the formula for probabilities expressed as the difference of utilities, the probabilities for the first individual are:

$$P_{ni} = \frac{1}{1 + \exp\left[-\left(V_{ni} - V_{nj}\right)\right]}$$

$$P_{11} = \frac{1}{1 + \exp\left[-\left(-2.8 - (-2.9)\right)\right]} = 52.5\%$$

$$P_{12} = \frac{1}{1 + \exp\left[-\left(-2.9 - (-2.8)\right)\right]} = 47.5\%$$

Using the alternative formula, the probabilities for the second individual are:

$$P_{ni} = \frac{\exp\left(V_{ni}\right)}{\exp\left(V_{ni}\right) + \exp\left(V_{nj}\right)}$$

$$P_{21} = \frac{\exp\left(-2.9\right)}{\exp\left(-2.9\right) + \exp\left(-3.0\right)} = 52.5\%$$

Table 2.2 Utility calculations for two individuals

Choice scenario for first individual			
Itinerary	Price ($100s)	Stops	$V_i = -0.4(Price_i) - 0.5(Stops_i)$
1	$7	0	$V_1 = -0.4(7) - 0.5(0) = -2.8$
2	$6	1	$V_2 = -0.4(6) - 0.5(1) = -2.9$
Choice scenario for second individual			
Itinerary	Price	Stops	$V_i = -0.4(Price_i) - 0.5(Stops_i)$
1	$6	1	$V_1 = -0.4(6) - 0.5(1) = -2.9$
2	$5	2	$V_2 = -0.4(5) - 0.5(2) = -3.0$

$$P_{22} = \frac{\exp(-3.0)}{\exp(-2.9) + \exp(-3.0)} = 47.5\%$$

The fact that only differences in utility are uniquely identified also influences the specification of error terms. Specifically, the following utility equation:

$$U_{ni}^1 = V_{ni} + \varepsilon_{ni},\ \varepsilon_i \sim G(\eta_i, \gamma)$$

is equivalent to the following model that adds a constant, η_i, to the systematic portion of utility and subtracts a constant, η_i, to the location parameter of the Gumbel distribution:

$$U_{ni}^2 = (V_{ni} + \eta_i) + (\varepsilon_{ni} - \eta_i),\ \varepsilon_i \sim G(0, \gamma)$$

Thus, the mode of the Gumbel distribution associated with each alternative must be set to a constant. A common normalization is to assume the mode is zero.

Specification of Alternative-specific Variables

The fact that only differences in utility are uniquely identified influences the way in which socio-demographic and other variables that do not vary across the choice set must be included in the utility function. Variables can be classified as generic or alternative-specific. Variables such as the price and stops variables shown in the itinerary choice scenario in Table 2.2 are "generic" because they can take on different values within an individual's choice set. In contrast, variables like an individual's annual income take on a "specific" value within that individual's choice set. Because only differences in utilities are uniquely identified, variables that do not vary across choice sets must be interacted with a generic variable or made "alternative-specific." In addition, given J alternatives, at most $J - 1$ alternative-specific variables can be included in the utility functions.

These concepts are illustrated in Table 2.3, which adds an additional variable, income, to the itinerary choice scenario. The need to specify income as alternative-specific variable or interact it with a generic variable can be easily seen with data in the idcase-idalt format, where each row represents a unique observation (or case) and alternative. Utility equations for alternatives one and two are defined as:

$$V_i = \beta_1 Cost_i + \beta_2 Stops_i + \beta_3 Income, \quad i = \text{first alternative}$$

$$V_j = \beta_1 Cost_j + \beta_2 Stops_j + \beta_4 Income, \quad j = \text{second alternative}$$

Further, since only differences in utility are identified and income does not vary across the choice set, only the difference $\beta_3 - \beta_4$ is uniquely identified. It is common to normalize the model by setting one of these parameters to zero. Setting alternative two as the reference alternative is equivalent to stating that $\beta_4 = 0$. The income coefficient, β_3 represents the effect of higher incomes on the probability of choosing alternative one (relative to the reference alternative). A negative (positive) value for β_3 would mean that individuals with higher incomes are less (more) likely to choose alternative one and more (less) likely to choose alternative two than individuals with lower incomes.

A second way to include income in the model is to interact it with a generic variable (*e.g.*, resulting in cost/income). The selection of the "best" generic variable should be motivated by behavioral hypotheses. Dividing cost by income reflects the analyst's hypothesis that high-priced itineraries are more onerous for individuals with lower incomes than for individuals with higher incomes. The utility equations in this case are defined as:

$$V_i = \beta_1 Stops_i + \beta_2 Cost_i / Income$$

$$V_j = \beta_1 Stops_j + \beta_2 Cost_j / Income$$

Table 2.3 Specification of generic and alternative-specific variables

IDCASE	IDALT	Cost ($)	Stops	Income ($)	Cost/Income
1	1	700	0	40,000	0.0175
1	2	600	1	40,000	0.0150
2	1	600	1	60,000	0.0100
2	2	500	2	60,000	0.0083

A second example based on no show data from a major U.S. airline is shown in Table 2.4. The analysis uses data for inbound itineraries departing in continental U.S. markets in March 2001. The results shown in Table 2.4 are for inbound itineraries (one-way itineraries are excluded from the analysis) and are based on 1,773 observations. The attributes shown in Table 2.4 are all categorical. Thus, when specifying the utility function, one of the categories is set to zero. That is, given *N* categories, at most *N-1* can be included in the utility equation. This is because given information about *N-1* categories, the value of the reference category is automatically known. For example, if we know that the passenger is traveling alone, we automatically know the passenger is not traveling in a group. Including all *N* categories in the model creates a situation in which there is perfect correlation, and the model cannot be estimated. Similar logic applies to why one of the alternatives must be set as a reference alternative. The parameters associated with the categories included in the utility function provide information on how much more likely (for β's > 0) or less likely (for β's < 0) the alternative is chosen compared to the reference category. In practice, the alternative that is chosen most often, the alternative that is available in the majority of choice sets, and/or the alternative that makes the interpretation of the β's easiest is used as the reference category.

Parameter estimates and t-stats shown at the bottom of Table 2.4 indicate that passengers with e-tickets are much more likely to show than passengers who do not have e-tickets. E-ticket is a very powerful predictor of no show rates because it helps discriminate among speculative and confirmed bookings; bookings that are not e-tickets have either not been paid for or have been paid for and confirmed via another purchase medium like paper tickets. Those traveling with another person (on the same booking reservation) and those who are general members of the carrier's frequent flyer program are also more likely to show. More interesting, booking class, one of the key variables used to predict no show rates in many current airline models, is not significant at the 0.05 level.

Specification and Interpretation of Alternative-specific Constants

Alternative specific constants (ASCs) are often included in utility functions. An ASC is similar to the intercept term used in linear regression and captures the average effect of all unobserved factors left out of the model. The inclusion

Table 2.4 Specification of categorical variables for no show model

IDCASE	IDALT	First/ Bus	High Yield	Low Yield	No FF	Gen FF	Elite FF	E-tkt	No E-tkt	Grp of 2+	Travel alone
1	1 SH	0	1	0	0	0	1	1	0	0	1
1	2 NS	0	1	0	0	0	1	1	0	0	1
2	1 SH	0	0	1	1	0	0	0	1	1	0
2	2 NS	0	0	1	1	0	0	0	1	1	0
3	1 SH	1	0	0	0	1	0	1	0	0	1
3	2 NS	1	0	0	0	1	0	1	0	0	1
4	1 SH	0	0	1	0	1	0	1	0	1	0
4	2 NS	0	0	1	0	1	0	1	0	1	0

include 2 include 2 include 1 include 1

OBS	ALT	Constant	First/ Bus	High Yield	Gen FF	Elite FF	E-tkt	Grp of 2+
1	1 SH	1	0	1	0	1	1	0
1	2 NS	1	0	1	0	1	1	0
2	1 SH	1	0	0	0	0	0	1
2	2 NS	1	0	0	0	0	0	1
3	1 SH	1	1	0	1	0	1	0
3	2 NS	1	1	0	1	0	1	0
4	1 SH	1	0	0	1	0	1	1
4	2 NS	1	0	0	1	0	1	1
$\boldsymbol{\beta}$	Show is ref.	1.40	-0.21	0.17	-0.56	0.05	-1.51	-0.47
t-stat		10.6	-1.1	1.4	-4.4	0.3	-13.0	-3.8
Sig		<0.001	0.128	0.074	<0.001	0.371	<0.001	<0.001
OR		4.07 (0.78)	0.81	1.19	0.57	1.05	0.22	0.63

Note: results from choice-based sample. The odds ratio (OR) for constant shown in parenthesis has been adjusted to reflect population shares.

of alternative specific constants in a model can be emphasized by using α's to represent the parameter estimates associated with alternative specific constants and $\boldsymbol{\beta}$'s to represent other parameter estimates:

$$V_i = \alpha_i + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_K X_{Ki}$$

By definition, a constant does not vary across a choice set, so constants must be specified as "alternative specific" variables. The same normalization rules discussed earlier apply, i.e., given J alternatives, at most $J-1$ (non-stratified) constants can be included in the model.

When validating the predictive performance of a binary or multinomial logit model, it is important to remember that, by including alternative specific constants in the model, the average probabilities obtained from applying the model to the N observations in the estimation dataset will closely approximate the choice probabilities in the estimation dataset. Thus, one cannot evaluate the predictive performance of the model by measuring how closely the model reproduces sample shares. In situations where sample shares do not represent population shares, the inclusion of a full set of identifiable ASCs in a multinomial logit model provides the analyst with a way to reproduce population shares. Specifically, the constants in a binary or multinomial logit model can be adjusted using the following relationship:

$$\alpha_i^{Pop} = \alpha_i^{sample} - \ln\left(H_i / Q_i\right)$$

where H_i is the sample share of alternative i and Q_i is the population share. As an example, the no show data shown in Table 2.4 are from a choice-based sample. That is, because the carrier's actual data contain millions of monthly booking transactions, a choice-based sample based on individual bookings was selected with approximately equal choice frequencies for the show and no show alternatives. Population choice probabilities are 89.6 percent (show) and 10.4 percent (no show), whereas sample choice probabilities are 45.9 percent (show) and 54.1 percent (no show). Thus, in the no show model of Table 2.4, the no show constant can be adjusted as:

$$\alpha_{NS}^{Pop} = 1.40 - \ln\left(0.541/0.104\right) = -0.249$$

The proof of this relationship is attributed to McFadden (not published but reported in Manski and Lerman 1977). Further, this relationship applies only to binary logit and multinomial logit models. Constant adjustments for a limited set of special cases for more complex models (including the nested logit and generalized nested logit models discussed in Chapters 3 and 4, respectively) can be derived by transforming these models into their equivalent Network Generalized Extreme Value model (discussed in Chapter 5). See Bierlaire, Bolduc and McFadden (2008) for the proof.

Conceptually, by adjusting ASCs, the analyst is changing only the intercept —not the tradeoffs represented in the $\boldsymbol{\beta}$'s—in order to match population shares. Stratified constants enable the analyst to match population shares along two dimensions (e.g., choice frequency by income group). That is, instead of defining a single constant for each alternative, multiple constants (one for each income group), are associated with each alternative. This is common in mode choice models as it enables the analyst to "match" observed population shares for each income group by adjusting the income-specific ASCs.

However, it is not always possible to include ASCs for every alternative. This occurs in situations in which the universal choice set is very large, such as in choice-based revenue management models. The universal choice for major carriers often contains dozens of alternatives, each representing a unique product sold for a specific itinerary. These products are defined by price and one or more ticket restrictions (e.g., advance purchase, Saturday night stay, minimum stay, and refundability and exchange criteria). In this type of choice scenario, it is not viable to define dozens of constants specific to each product. In this case, constants can be omitted and/or grouped into meaningful categories. For example, in departure time models used for urban travel demand modeling applications, the constants for infrequently chosen alternatives across multiple adjacent departure times can be combined into a single constant.

Odds Ratios

Odds ratios are frequently used to interpret the coefficients of logistic regression and binary logit models. They can also be used with more complex models, including the multinomial logit model. Conceptually, the binary logit model and logistic regression are similar. Both models predict the probability that one out of two discrete alternatives will be chosen. In logisitic regression, the response variable, y, is defined to be the log of the odds, where odds is defined as the ratio of probabilities for two alternatives. Given that $P_1 = 0.75$ and $P_2 = 0.25$, the "odds of P_1" is $0.75/0.25 = 3$, i.e., alternative one is three times more likely to be selected than alternative two. Similarly, the "odds of P_2" is $0.25/0.75 = 0.33$.

Often, the analyst is interested in knowing how the odds differ across an observed categorical variable. As an example, assume that one out of five business passengers no show for their flights, whereas one out of 20 leisure passengers no show. The odds ratio provides information on how much more likely business passengers are to no show compared to leisure passengers. Specifically, $P_{NS} \mid Business = 1/5 = 0.20$ and $P_{NS} \mid Leisure = 1/20 = 0.05$. The ratio of business travelers and leisure travelers no show rates, $0.20/0.05 = 4.0$, means that business passengers are four times more likely to no show than leisure passengers. This example frames the interpretation of the odds ratio in terms of a variable with two categories (i.e., the passenger is either traveling for business or leisure). However, the interpretation of the odds ratio can be generalized to include multiple categories and continuous variables.

Formally, noting that $V_i = \alpha_i + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_K X_{Ki}$, the logistic regression equation for the log of the odds of P_1 with alternative two as the reference is given as:

$$\log\left(\frac{P_1}{P_2}\right) = \log\left(\frac{\dfrac{e^{V_1}}{e^{V_1}+e^{V_2}}}{\dfrac{e^{V_2}}{e^{V_1}+e^{V_2}}}\right) = \log\left(e^{V_1 - V_2}\right) = V_1 - V_2$$

$$= \alpha + \beta_1\left(X_{11} - X_{21}\right) + \beta_2\left(X_{12} - X_{22}\right) + \ldots + \beta_K\left(X_{1K} - X_{2K}\right)$$

In the case where no generic variables are included in the utility function and alternative two is set as the reference category, the log of the odds of P_1 reduces to the following (which is the formula more commonly shown in a logistic regression context):

$$\log\left(\frac{P_1}{P_2}\right) = \alpha + \beta_1 X_{11} + \beta_2 X_{12} + \ldots \beta_K X_{1K}$$

The odds of P_1 are obtained by taking the exponent of each side:

$$\frac{P_1}{P_2} = \exp^{\alpha+\beta_1 X_{11}+\beta_2 X_{12}+\cdots+\beta_K X_{1K}}$$

The odds ratio is defined as the change in the logs of the odds due to a "unit change" in x_{1k}, holding all other variables constant. In the context of categorical variables, the "unit change" reflects increasing the value of x_{1k} from zero to one. For example, in the no show model shown in Table 2.4, the odds ratio associated with groups is 0.63. This means that those traveling in a group are 0.63 times less likely to no show (or 1/0.63=1.6 times more likely to show) than those traveling alone. Similarly, the odds associated with frequent flyer status indicate that, compared to the reference category (non-frequent flyer members), general members are 0.57 times less likely to no show, whereas elite members are slightly more likely to no show, all other factors being held constant.

The same definition of the odds ratio applies to continuous variables, i.e., the odds ratio reflects the change in the logs of the odds due to a "small" change in x_{1k}, holding all other variables constant. As discussed in Long (1997), the value associated with the change, δ, can be defined in different ways. A unit change is defined when $\delta = 1$. However, this measure can be sensitive to the units of measurement (e.g., a one unit change to a variable measured in minutes may provide a different interpretation than a one unit change to a variable measured in hours). To standardize changes across different units of measurement, δ can be defined to represent the standard deviation of x_k. Formally, the odds ratio is obtained by comparing the odds of $\delta = 0$ to the odds when $\delta > 0$:

$$OR = \frac{\exp^{\alpha+\beta_1 X_{11}+\ldots+\beta_k(X_{1k}+\delta)+\cdots+\beta_K X_{1K}}}{\exp^{\alpha+\beta_1 X_{11}+\ldots+\beta_k X_{1k}+\cdots+\beta_K X_{1K}}} = \exp(\beta_k \delta)$$

$$= \exp(\beta_k) \text{ for } \delta = 1 \text{ and for categorical variables.}$$

From an interpretation perspective, it is important to note that the relationship between the odds ratios and predicted probabilities is not linear. This means that doubling the odds of P1 does not correspond to doubling the probability that P1 will occur. For example, if the odds are very small (1/100) and doubled (1/50),

the corresponding probability of P1 will remain small. To visually examine the relationship between the odds and predicted probabilities, it is common to use enhanced odds ratio plots (Long 1997; Long and Freese 2003; StataCorp 2008), particularly when using models with multiple outcomes. For example, an odds ratio of two given by (1/100):(1/50) will be shown to have a relatively small impact on the predicted probabilities on an enhanced odds ratio plot, whereas an odds ratio of two given by (1/4):(1/2) will be shown to have a relatively large impact on an enhanced odds ratio plot. Formally, in an enhanced odds ratio plot, the height of the letters is proportional to the square root of the discrete change in the odds. The lack of significance between two categories is noted by a dotted line.

Figure 2.11 illustrates the odds ratio plot and enhanced odds ratio plot for the no show model reported in Table 2.4. An enhanced odds ratio plot provides information on "how much" different the predicted probabilities associated with the show and no show choices are relative to a reference point. The reference point for categorical variables is defined as the reference category and "reasonable values" for continuous variables. All of the variables in the no show model are categorical. As an example, consider e-ticket, which is represented in the dataset as an indicator variable *e-ticket* equal to one if the individual purchased a ticket electronically and zero otherwise. Figure 2.11 shows that *e-tickets* are "much less likely" to no show, all other factors being held constant. Here, "much" is represented by the height of the letter 2 identifying the no show category and "less likely" is represented by the fact that 2 is underlined. The height of the letter relates to the discrete changes in probabilities and dotted lines between categories indicate that the difference in discrete probabilities is not statistically different at the 0.05 level.

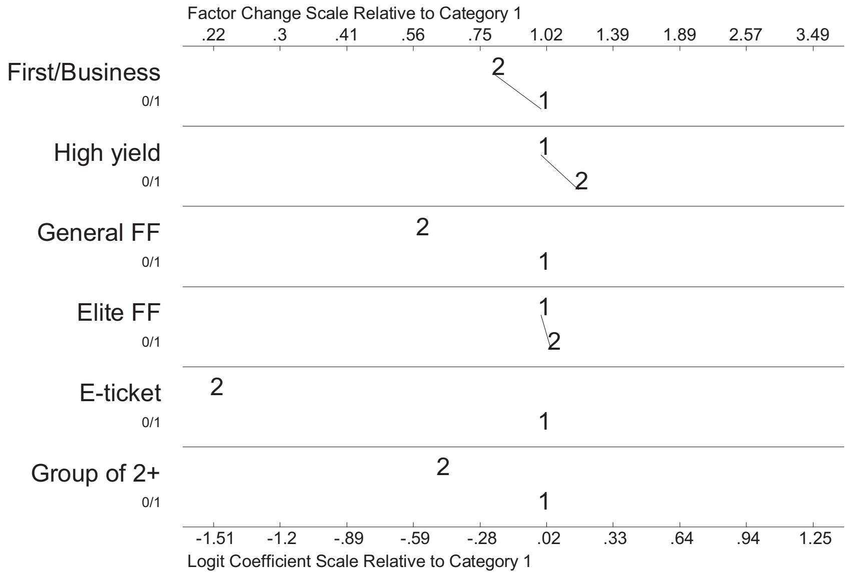
Interpretation of the Scale Parameter

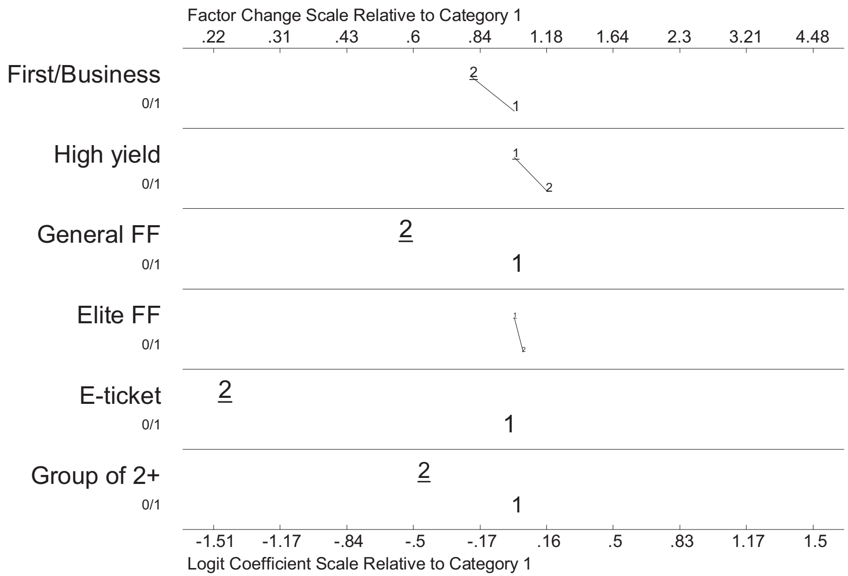
The interpretation of $\boldsymbol{\beta}$'s is also linked to the scale parameter. The binary logit probabilities were derived under the assumption that error terms were independently and identically Gumbel distributed with mode zero and scale one. The assumption that the mode of the distribution was centered at zero is not restrictive, as adding a constant value to each utility does not affect which alternative has the highest utility. Similarly, multiplying each utility by a constant will not affect which alternative has the highest utility. Formally, the following two utility expressions:

$$U_{ni}^{1} = V_{ni} + \varepsilon_{ni}$$

$$U_{ni}^{2} = \zeta V_{ni} + \zeta \varepsilon_{ni}, \quad \zeta > 0$$

are equivalent in the sense that the alternative with the maximum utility given by the first utility equation, U_{ni}^{1}, is the same as the alternative with the maximum utility given by the second utility equation, U_{ni}^{2}, which multiples the utility of each





Note: Category 1 is Show, Category 2 is No Show

Figure 2.11 Odds ratio and enhanced odds ratio plots for no show model

alternative by the constant ζ. Note that, consistent with the earlier definition, the scale parameter is defined according to the "inverse variance" relationship, thus when $\zeta = 3$, the variance is $\pi^2 / (6 \times 3^2)$.

However, although adding a constant value to each utility was shown not to affect choice probabilities, multiplying each utility by a constant value will affect choice probabilities. If the scale ζ is large (e.g., utility is measured with very little error), then ζV_{ni}'s are large and the differences in utility ζV_{ni}-ζV_{ni} are large. Thus, the probabilities will be more extreme (closer to zero or one) and the S-shape curve will be more steep. Given high (low) uncertainty about utility, choices can be predicted with less (greater) certainty. The relationship between the scale and choice probability is shown in Figure 2.12.

The scale parameter also influences the interpretation of the $\boldsymbol{\beta}$ parameter estimates, because the estimate of $\boldsymbol{\beta}$ cannot be identified separately from the scale parameter. This point can be made explicit by considering the case where the scale is not normalized to one. Assume the "true" utility for alternative i is given as:

$$U_i = \sum_k X_{ik} \beta_k + \varepsilon_i \quad \varepsilon_i \sim G(0, \gamma)$$

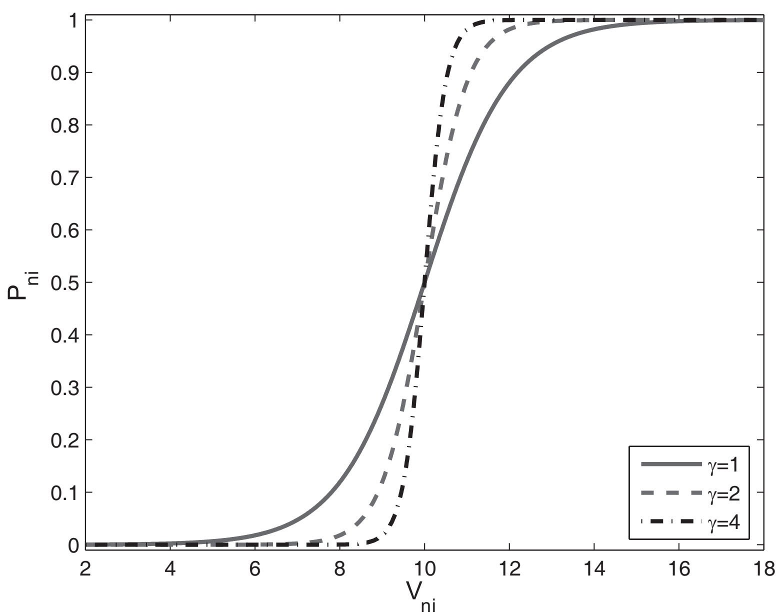


Figure 2.12 Relationship between binary logit probabilities and scale

Source: Adapted from Koppelman 2004: Figure 2.6 (reproduced with permission of author).

Defining $U_i^* = U_i \times \gamma$, the utility for alternative i becomes:

$$U_i^* = \sum_k X_{ik}\beta_k\gamma + \varepsilon_j\gamma \quad \left(\varepsilon_j\gamma\right) \sim G(0,1)$$

That is, the parameter estimates are actually the "true" $\boldsymbol{\beta}$ parameters multiplied by the scale of utility. Thus, $\boldsymbol{\beta}$ parameter estimates will be larger when the scale γ is large (i.e., the variance is small). This is one reason why the absolute magnitude of parameter estimates cannot be compared across models that come from different data sources. Instead, other measures must be used to interpret the $\boldsymbol{\beta}$ parameters. For example, the interpretation of the ratio of two parameter estimates, $(\beta_1\gamma)/(\beta_2\gamma)$, is not influenced by the scale parameter.

Finally, it is important to note that while the scale of utility in the binary logit model is normalized by setting the scale to one, more complex normalizations may be required for models such as the probit and mixed logit in which the scale (or variance) is allowed to differ across alternatives. An example of the normalization procedure for mixed logit models is presented in Chapter 6.

Multinomial Logit Model

Choice Probabilities

The multinomial logit (MNL) model is a generalization of the binary logit model and is used to describe how an individual chooses among three or more discrete alternatives. As with the binary logit model, MNL probabilities are derived from the assumption that error terms are distributed Gumbel with mode zero and scale one (which implies a variance of $\pi^2/6$). The MNL probabilities (derived in Appendix 2.1 at the end of the chapter) are given as:

$$P_{ni} = \frac{\exp\left(V_{ni}\right)}{\sum_{j \in C_n} \exp\left(V_{nj}\right)}$$

An alternative formula for the MNL probability, which more clearly shows that only differences in utility are identified, is obtained by dividing the numerator and denominator by exp (V_{ni}), or

$$P_{ni} = \frac{1}{\sum_{j \in C_n} \exp\left[-\left(V_{ni} - V_{nj}\right)\right]}$$

The same concepts discussed in the context of binary logit models also apply to the interpretation and specification of MNL models. The first concept relates to the fact that only differences in utility are uniquely identified. This property

requires that variables that do not vary over the choice set be included in the utility function either by interacting them with a generic variable or by specifying them as alternative-specific (with the parameter associated with one alternative normalized to a constant value, i.e., zero). This property also requires that the location parameter associated with the Gumbel distribution be normalized to a constant, i.e., zero.

The second concept relates to the fact that the alternative with the largest utility is unaffected by the scale of utility. This property influences the interpretation of $\boldsymbol{\beta}$ parameter estimates and choice probabilities in several ways. First, for identification purposes, the scale must be normalized to a constant, i.e., one. This influences the interpretation of $\boldsymbol{\beta}$ parameter estimates in the sense that the magnitude of these parameters is influenced by the amount of variance in the model. The $\boldsymbol{\beta}$ parameter estimates will be lower in situations where the variance associated with unobserved factors is high. This is one reason why the absolute magnitude of parameter estimates cannot be compared across different datasets. To remove the scale effect, odds ratios or ratios of parameters (such as value of time calculations) are used to interpret parameter estimates across different datasets. Second, it is important to note that although the alternative with the largest utility is unaffected by the scale of utility, choice probabilities are affected. Probabilities can be predicted with more precision when the variance associated with the unobserved factors is smaller. Conceptually, this can be thought of in terms of a "signal-to-noise" ratio. The stronger the "signal" (reflected in the observed portion of utility) compared to the "noise" (reflected in the variance of the unobserved portion of utility), the steeper the logit probability curve will be. In the extreme case where variance is zero, choice probabilities become deterministic.

The final concept introduced in the context of the binary logit model that also applies to the MNL model relates to prediction. Specifically, it is important to remember that when a full set of identified alternative specific constants are included in a binary logit or MNL model, the model will be able to replicate sample shares.

In addition to these three fundamental concepts, it is common to use direct- and cross-elasticities to examine and understand the substitution patterns of MNL and more complex discrete choice models. Other metrics related to forecasting performance of models are particularly relevant in the airline industry, where even slight improvements in accuracy can translate to millions of dollars in incremental revenue.

Direct- and Cross-elasticities

Although odds ratios are used to interpret the sensitivity of the log of the odds to a unit change in one of the observed factors, derivatives of choice probabilities directly capture the sensitivity of choice probabilities to a unit change in one of the observed factors (all other factors being held constant). The sensitivity of P_{ni} with respect to a change in the k^{th} variable associated with alternative i, x_{ik}, is often referred to as the "direct effect," as it represents an associated change in market

share for alternative i due to making a change "directly" to the characteristics of alternative i. In addition, it is also useful to examine how a change in x_{ik} affects the sensitivity of P_{nj}. This is often referred to as the "cross effect." Consistent with the discussion related to odds ratios and enhanced odds ratios, it is important to remember that, as with all derivatives, the sensitivity of probabilities is measured at a specific point and may change depending on the value of x_{ik}.

By definition, derivatives capture the effect of a unit change in one variable, and are thus sensitive to the units of measurement (e.g., is the unit of measurement associated with travel expressed in seconds, minutes, or hours?). Elasticities are often used in place of derivatives, as they control for the units of measurement. Formally, suppressing the index n associated with the individual, the elasticity of P_i with respect to a percentage change in the k^{th} attribute for alternative i, x_{ik}, is defined as:

$$\eta_{X_{ik}}^{P_i} = \frac{\partial P_i}{\partial X_{ik}} \cdot \frac{X_{ik}}{P_i}$$

Similarly, the cross-elasticity of P_j with respect to a percentage change in the k^{th} attribute for alternative i, x_{ik}, is defined as:

$$\eta_{X_{ik}}^{P_j} = \frac{\partial P_j}{\partial X_{ik}} \cdot \frac{X_{ik}}{P_j}$$

Appendices 2.2 and 2.3 show the direct-elasticity and cross-elasticity derivations for the MNL model:

$$\eta_{X_{ik}}^{P_i} = \left(1 - P_i\right)\beta_k X_{ik}$$

$$\eta_{X_{ik}}^{P_j} = -P_i \beta_k X_{ik}$$

It is important to note that cross-elasticities are equal for all $j \neq i$. That is, the percentage changes in probabilities associated with a percentage change in x_{ik} are identical. This is due to the underlying "independence of irrelevant alternatives" or IIA property of the MNL model.

Independence of Irrelevant Alternatives

The IIA property of the MNL model states that the ratio of choice probabilities between any two alternatives is independent of the availability or attributes of the other alternatives. Formally, given alternatives i and k:

$$\frac{P_i}{P_k} = \frac{e^{V_i} \Big/ \sum_{j \in C_n} e^{V_j}}{e^{V_k} \Big/ \sum_{j \in C_n} e^{V_j}} = \frac{e^{V_i}}{e^{V_k}} = e^{V_i - V_k}$$

That is, the ratio of probabilities for alternatives *i* and *k* depends only on the utilities for those alternatives. The IIA property is most commonly referenced in the context of the behavioral limitations it imposes on the choice situation. One colloquial expression often used to refer to this limitation is the "red bus, blue bus" problem. Specifically, consider a situation in which an individual is choosing between taking a red bus and driving to work. The probabilities the individual will drive or take the red bus are 0.75 and 0.25, respectfully. The city decides to put a second bus on the route. This bus is painted blue but is identical in all aspects to the current red bus (e.g., same stops, same interior, etc.). If the blue bus is added in as a third alternative, the MNL model will predict that share be drawn proportionately from the other two alternatives, with resulting probabilities of 0.60, 0.20, and 0.20 for the drive, red bus, and blue bus alternatives. Note the ratio of choice probabilities for drive and red bus 3:1 is the same in both scenarios. Intuitively, however, the analyst does not expect that the introduction of a second bus will draw share from the drive alternative. Clearly, the example is contrived in the sense that the analyst would never define a new alternative that is identical to another alternative except for a variable that is irrelevant to the choice dimension. However, the example illustrates one of the main limitations of the MNL model (i.e., the IIA property). Specifically, by imposing assumptions on the distribution of error terms representing factors left out of the model, the analyst is also imposing assumptions on substitution patterns among alternatives. This is one reason why more advanced models that relax these assumptions report the direct- and cross-elasticities. Direct- and cross-elasticities provide insight into the relationship between substitution patterns and how changes in observed variables or the addition of a new alternative draws share from the other alternatives.

Table 2.5 Example of the IIA property

	Drive	**Red Bus**	**Blue Bus**
Probability for two alternatives	0.75	0.25	N/A
MNL probability for three alternatives	0.60	0.20	0.20
"Expected" probability for three alternatives	0.75	0.125	0.125

While one of the main limitations of MNL models is often quoted to be the IIA property, it is important to note that in many practical applications, this property is quite useful. Specifically, this property is often leveraged in two ways. The first involves situations, such as the one described above, in which alternatives are added or dropped from the choice set. For example, Ratliff, Venkateshwara, Narayan, and Yellepeddi (2008) use a MNL model to predict the probability an airline itinerary will be selected and use the IIA property to calculate recapture rates. Recapture rates are used to redistribute passengers to other itineraries when

one itinerary becomes unavailable (due to reaching capacity). From the perspective of the airline whose itinerary is no longer available, recapture rates distinguish between those passengers who are "recaptured" on its other itineraries versus those who are "captured" by other airlines. Ratliff (2006) also applies the IIA property in other revenue management contexts, including upsell/downsell and unconstrained demand. Although the IIA property is likely violated for itinerary choice applications, Ratliff's methodology nonetheless represents a substantial improvement over recapture rate methods currently used in practice. Extensions to more advanced logit specifications, as well as consideration of outbound and inbound differences in recapture rates, are all valuable research extensions to work that has been done in this area.

The second way in which the IIA property is often leveraged is when dealing with datasets that have large choice sets, such as destination choice models used in urban travel demand models. Assuming the IIA property holds, the ratio of choice probabilities between any two alternatives is irrelevant of the availability of alternatives, which means that it is possible to exclude some alternatives from the choice set (through sampling) while still obtaining consistent parameter estimates. The sampled choice set contains the chosen alternative in addition to the sampled alternatives.

Using a sample of alternatives to estimate the parameters of a MNL model was one of the first techniques proposed to formally test the appropriateness of the IIA property (McFadden 1978). Conceptually, if IIA holds, then the parameter estimates obtained from the sample should not be significantly different from the parameter estimates obtained from the full dataset. Several other tests of IIA are discussed in Train (2003: 53–4). However, in practical applications, these tests are somewhat limited because although they can detect the violations of the IIA property, they do not provide guidance as to whether the violation can be overcome by using a different specification of the observed portion of utility and/or whether other models that relax the IIA property are more appropriate. Consequently, in practice, it is common for the analyst to first develop a well-specified utility function and then test relaxations of assumptions on the error components (e.g., by estimating nested logit, generalized nested logit, mixed logit, etc.) and see which model fits the data the best. Formal tests comparing the MNL to NL and other discrete choice models are also used to help guide the analyst in the selection of a preferred model. The modeling process, along with these tests, is covered in depth in Chapter 7.

Evaluation of Forecasting Performance

One important part of modeling individual behavior focuses on developing a well-specified utility function that captures how individuals make trade-offs among different variables. Formal statistical tests help guide the selection of variables, variable forms, and specific models (such as the MNL, nested logit, mixed logit, etc). Also, depending on the goals of the research and/or research field, it may be common to report results using odds ratios, or elasticities and cross-elasticities.

Depending on the research context, several other measures, such as those tied to consumer surplus or compensating variation may also be relevant to interpreting the results of the discrete choice model.

In addition to understanding the behavioral interpretations and substitution patterns of discrete choice models, it is important to evaluate the forecasting accuracy. Forecasting accuracy tends to receive a greater priority in airline research than in urban travel demand, marketing, and other research areas that have traditionally used discrete choice models. This is because the stakes are high in the airline industry—even small improvements (or deteriorations) in the forecasting accuracy of key decision support models may lead to millions of dollars of revenue gains (or losses) for an airline. For example, Stefan Polt, affiliated with Lufthansa German Airlines, stated that "as a rule of thumb, a 10 percent improvement in (demand) forecasting accuracy translates to a 1–2 percent revenue increase" (Polt 2002).

Given that many applications using discrete choice models naturally lend themselves to the ability to perfectly replicate samples, validation, and population shares via the adjustment of ASCs, how should the forecasting performance be measured? In the airline industry, the performance of a discrete choice model is often evaluated in terms of how it interacts with the entire decision support system, which typically is at a level of aggregation that is not the same as that represented in the constants. In the case of itinerary share models, this often involves measuring the accuracy of itinerary choices at a different level of aggregation, namely for flight legs (which represent the level at which business decisions of where and when to schedule a flight are made). The forecasting performance of the discrete choice model is also typically benchmarked against the forecasting model currently being used.

The evaluation of forecasting performance in the airline industry tends to be different than those used in urban travel demand or marketing applications. For example, in urban travel demand applications, discrete choice models support the evaluation of the benefits and costs associated with long-term transportation infrastructure improvements and/or new demand management policies (such as the benefits in travel time and/or air pollution reductions due to adding a lane to a highway or imposing time-of-day tolls). Often, when evaluating forecasting performance in these applications, the relative costs and benefits associated with different scenarios are more important than the absolute values. Further, the ability to evaluate the forecasting performance of travel demand models is often limited due to the time lag between (potential) implementation of infrastructure improvements and subsequent shifts in traveler behavior. However, in the airline industry, the ability to evaluate forecasting performance of different models is easier and almost "instantaneous." This is because airline forecasting models form the backbone of many core decision support systems that span revenue management, scheduling, and other areas. Further, the decisions these models help support, such as where to schedule flights in a market or how many seats to sell on a flight, are taken on a quarterly, weekly, or even daily basis. For these reasons,

it is the author's opinion that the forecasting methods typically discussed in the context of discrete choice models for travel demand applications (such as the use of synthetic populations or complete sample enumeration, and concerns related to the introduction of forecasting bias when using the average values) are less relevant in airline applications.

Estimation Methods

Given an understanding of the basic properties of discrete choice models required to specify and interpret models, this section focuses on the underlying methodology and concepts used to solve for the parameters of the choice models. For all choice problems, an observation represents a decision-maker, a vector of attributes associated with the decision-maker and alternatives, and the chosen alternative. The problem of interest is to solve for the parameters $\boldsymbol{\beta}^*$ given a random sample of observations (extensions to other types of samples are discussed later in this section). Estimators based on maximum likelihood estimation are most commonly used. Although other, more complex estimators can be used such as those based on the method of moments or the method of scores (e.g., see Train 2003), the focus of this discussion is on maximum likelihood estimators. Maximum likelihood estimation solves for the values of $\boldsymbol{\beta}$ that maximize the likelihood function:

$$L(\boldsymbol{\beta}) = \prod_{n=1}^{N} \prod_{i \in C_n} P\left(i \mid x_{ni,} \boldsymbol{\beta}\right)^{d_{ni}}$$

where:

N is the number of individuals in the random sample,

$i \in C_n$ are alternatives in the choice set C for individual n,

x_{ni} is the vector of attributes associated with alternative i and individual n,

d_{ni} is an indicator variable equal to 1 if individual n selects alternative i, and 0 otherwise,

$P(i \mid x_{ni}, \boldsymbol{\beta})$ is the probability of selecting alternative i given a sample of attributes x_{ni} and estimates $\boldsymbol{\beta}$. Earlier discussions defined this probability as P_{ni}. When the conditional form is used in this section, it is used to emphasize the fact that the probability is dependent on characteristics of the sampling distribution related to attributes and estimates.

Computationally, it is easier to maximize the logarithm of the likelihood function, i.e., the log likelihood (LL) function, or:

$$LL = \sum_{n=1}^{N} \sum_{i \in C_n} d_{ni} \ln P_{ni}$$

An example log likelihood calculation is shown in Table 2.6. The example is presented in the idcase-idalt format, where each row represents a unique

Table 2.6 Example of a MNL log likelihood calculation

OBS	ALT	d_{ni}	ASC train	ASC bus	Cost ($)	Time (hr)	Male Train	Male Bus	V_{ni}	P_{ni}	$d_{ni} \times \ln(P_{ni})$
1	1 Car	1	0	0	$4.00	0.75	0	0	-1.33	0.614	-0.487
1	2 Train	0	1	0	$3.00	0.50	1	0	-2.15	0.269	0
1	3 Bus	0	0	1	$1.75	1.00	0	1	-2.99	0.117	0
2	1 Car	0	0	0	$7.00	1.25	0	0	-2.23	0.331	0
2	2 Train	1	1	0	$5.50	0.33	0	0	-1.52	0.669	-0.402
3	1 Car	0	0	0	$5.00	1.00	0	0	-1.75	0.431	0
3	2 Train	0	1	0	$6.00	0.50	1	0	-2.30	0.249	0
3	3 Bus	1	0	1	$3.00	0.33	0	1	-2.05	0.321	-1.137
…											
$\boldsymbol{\beta}$			-0.75	-1.0	-0.05	-1.5	-0.5	-0.4			

observation (or case) and alternative. Note that in this example, separate columns are defined for each alternative-specific variable, e.g., the male variable appears in two columns: "Male Train" and "Male Bus." This is done to emphasize that the parameter for male associated with the train alternative (-0.5) applies only to those rows in which the individual is a male and the row represents the train alternative.

Conceptually, there are several subtle points related to the log likelihood function. First, the log likelihood associated with an observation is always negative, due to fact that P_{ni} falls between zero and one. Second, the quantity $d_{ni} \times \ln(P_{ni})$ will be closest to zero when the predicted probability associated with the chosen alternative approaches one. This represents the situation when alternative i was chosen by individual n and differences in utility between alternative i and all other alternatives in the choice set for individual n are large. Finally, note that observations that have choice sets that contain a single alternative may be eliminated from the estimation dataset, as these observations provide no information how the individual makes tradeoffs among two or more alternatives (i.e., the probability for this observation is known with certainty and is one).

The $\boldsymbol{\beta}$ parameter estimates are obtained by using optimization algorithms that maximize the log likelihood function. In the case of the binary logit and multinomial logit models, the log likelihood function is globally concave. This can be verified by examining its first and second derivatives with respect to $\boldsymbol{\beta}$. Given:

$$LL = \sum_{n=1}^{N} \sum_{i \in C_n} d_{ni} \ln \left(\frac{e^{V_{ni}}}{\sum_{j \in C_n} e^{V_{nj}}} \right)$$

$$LL = \sum_{n=1}^{N} \sum_{i \in C_n} d_{ni} \left(\beta' x_{ni} - \ln \sum_{j \in C_n} e^{V_{nj}} \right)$$

the derivative of the log likelihood function with respect to the k^{th} parameter is given as:

$$\frac{\partial LL}{\partial \beta_k} = \sum_{n=1}^{N} \sum_{i \in C_n} d_{ni} \left(x_{nik} - \frac{\sum_{j \in C_n} e^{\beta' x_{nj}} \cdot x_{njk}}{\sum_{j \in C_n} e^{\beta' x_{nj}}} \right), \quad k = 1, \ldots, K$$

Noting that d_{ni} is an indicator variable equal to 1 if individual n selects alternative i, and 0 otherwise, gives:

$$\frac{\partial LL}{\partial \beta_k} = \sum_{n=1}^{N} \sum_{i \in C_n} d_{ni} \left(x_{nik} - \frac{e^{\beta' x_{ni}} \cdot x_{nik}}{\sum_{j \in C_n} e^{\beta' x_{nj}}} \right), \quad k = 1, \ldots, K$$

$$\frac{\partial LL}{\partial \beta_k} = \sum_{n=1}^{N} \sum_{i \in C_n} \left(d_{ni} - P_{ni} \right) \cdot x_{nik}, \quad k = 1, \ldots, K$$

For reference, the second derivatives of the log likelihood function are:

$$\frac{\partial LL}{\partial \beta_k \partial \beta_l} = -\sum_{n=1}^{N} \sum_{i \in C_n} P_{ni} \cdot \left[x_{nik} - \sum_{j \in C_n} x_{njk} \cdot P_{nj} \right] \cdot \left[x_{nil} - \sum_{j \in C_n} x_{njl} \cdot P_{nj} \right]$$

The maximum of the log likelihood function is obtained when $\partial LL / \partial \boldsymbol{\beta} = 0$. Further, since the second derivative is negative semi-definite, the log likelihood function is globally concave (which implies there is one unique solution for $\boldsymbol{\beta}$ that maximizes the log likelihood function).

Although the log likelihood function for the binary logit and MNL is globally concave, the same is not true for more complex models, such as the nested logit and mixed logit models. Solution of these models requires non-linear optimization methods. Three of the most popular algorithms include the Newton-Rhapson method, BFGS, and BHHH. The BFGS algorithm is named after Broyden, Fletcher, Goldfarb, and Shanno and the BHHH algorithm is named after Berndt, Hall, Hall, and Hausman. Additional information on these algorithms can be found in Ruud (2000), Dennis and Schnabel (1996), and Nocedal and Wright (1999).

Interpretation of β Estimates Using Iso-utility Lines

Before extending the discussion of maximum likelihood estimators to other sampling designs, it is useful to provide a visual interpretation of the "optimal" $\boldsymbol{\beta}$ estimates using a simple example that uses the concept of iso-utility lines. Specifically, consider the choice between two alternatives (car and bus). The utility functions for these alternatives include two variables (time and cost). For simplicity, the alternative specific constants are suppressed. Specifically:

$$V_1 = \beta_1 Time_1 + \beta_2 Cost_1$$

$$V_2 = \beta_1 Time_2 + \beta_2 Cost_2$$

The value of time is obtained as the ratio of β_1/β_2. Note that the value of time is given as β_1/β_2, and not β_2/β_1. This is because the units associated with $\boldsymbol{\beta}$ are the inverse of the units associated with time and cost. Stated another way, because the utility function measures tradeoffs among attributes, utility is unitless. Thus, the units associated with the time and cost parameters are given as:

$$V[unitless] = (\beta_1) \cdot hr + (\beta_2) \cdot \$$$

$$V[unitless] = \left(\frac{1}{hr}\right) hr + \left(\frac{1}{\$}\right) \cdot \$$$

Figure 2.13 uses iso-utility lines to show the distinction between parameter estimates that indicate a high value of time and those that indicate a low value of time. Iso-utility lines are defined by a series of parallel lines where the value of time is represented by the inverse of the slope. The use of parallel lines is linked to the fact that only differences in utility are uniquely identified, i.e., the trade-off is independent of the absolute level of utility. The panel on the left of Figure 2.13 indicates a low value of time, i.e., the individual is not willing to spend a lot of money to save a "unit" of time. In contrast, the panel on the right of Figure 2.13 indicates a higher value of time because the individual is willing to spend more money to save the equivalent amount of time.

Iso-utility lines can also be used to understand the process used to find the values of $\boldsymbol{\beta}$ that fit the data the best. Figure 2.14 contains two choice scenarios: the first individual is faced with a choice between taking car 1 and bus 1 and chooses bus, which implies that the iso-utility line falls to the left of bus 1 and to the right of car 1. Similarly, given that the second individual chooses car, the slope of the iso-utility line will fall to the left of car 2 and to the right of bus 2. The iso-utility lines shown in the figure represent the range of slope parameters that "fit the data" in the sense that they will result in car 2 having a higher utility than bus 2 and bus 1 having a higher utility than car 1.

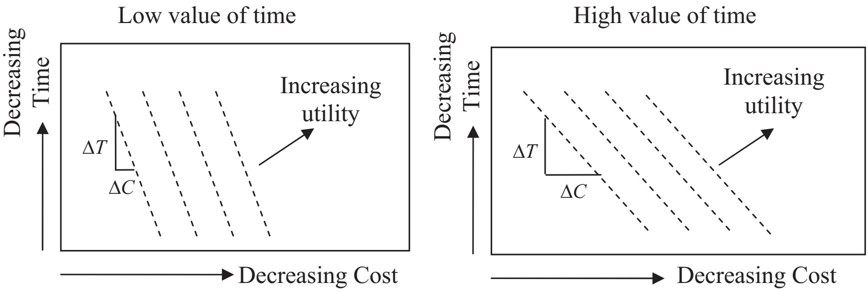


Figure 2.13 Iso-utility lines corresponding to different values of time

Source: Adapted from Koppelman and Bhat 2006: Figure 4.5 (reproduced with permission of authors).

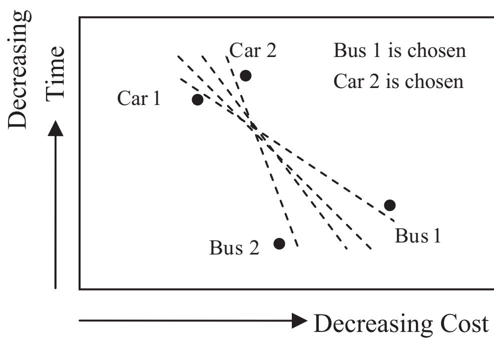


Figure 2.14 Interpretation of β using iso-utility lines for two observations

Source: Adapted from Koppelman and Bhat 2006: Figure 4.6 (reproduced with permission of authors).

The example is extended to multiple observations in Figure 2.15. Conceptually, the objective is to find slope parameters that result in the placement of diamonds to the right of the line (representing a correct prediction that alternative two is chosen) and circles to the left of the line (representing a correct prediction that alternative one is chosen). The utility function represented in Figure 2.15 has also been generalized to include an intercept term. If alternative-specific constants are excluded, the iso-utility line would intersect the *x*- and *y*-axes at zero.

Why Should Airlines Care About Estimation Based on Non-random Samples?

The maximum likelihood (ML) estimator is derived under the assumption that the estimation dataset is based on a random sample of observations from the population. However, different sample selection processes are often used when collecting data to ensure adequate representation of population groups and chosen alternatives. In

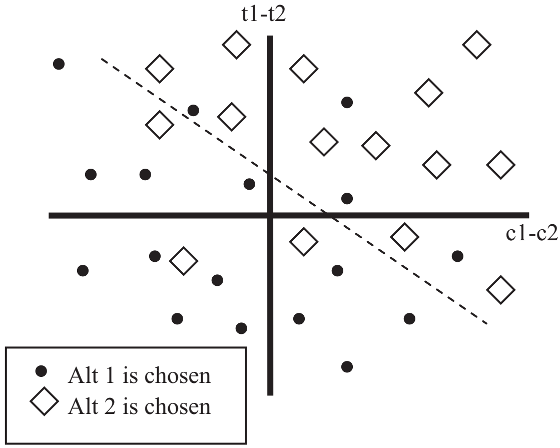


Figure 2.15 Interpretation of β using iso-utility lines for multiple observations

Source: Adapted from Koppelman 2004: Figure 2.7 (reproduced with permission of author).

general, sampling processes are characterized as random, exogenous, or choice-based (Manski and Lerman 1977). All individuals in a population have an equal probability of being selected in random samples; however, in exogenous (stratified) and choice-based samples, individuals are selected disproportionately according to observable attributes/segments and choice/purchase decisions, respectively.

Historically, in urban travel demand models, there is often a strong need to use non-random samples. For example, in many U.S. markets, travel surveys are typically collected every five to ten years by metropolitan planning organizations (MPOs) to support regional planning and infrastructure investment decisions. The evaluation of public transit options and an understanding of individuals' mode choice decisions are an important part of the regional planning process. In addition, it is particularly important to ensure these enhancements are equitable in the sense that both low income and high income individuals benefit from the portfolio of infrastructure improvements. However, compared to other countries, in the U.S. few individuals take public transit. In addition, survey response rates are typically lower among low income groups than high income groups. Thus, when collecting surveys about current travel behavior, it is desirable to over-sample low income groups and those currently choosing transit to ensure that there are sufficient observations (within the survey budget) to understand the behavior of these market segments.

In contrast to urban travel demand applications, the airline industry is fortunate to have numerous revealed preference data sources, many of which contain millions of passenger transactions. However, even in this industry, it is often advantageous to calibrate models using smaller random or non-random samples. One of the

"modeling myths" the author has often encountered when discussing statistical modeling with those in the airline industry is the perception that parameter estimates can only be accurate if the entire population data available to the analyst is used. However, the amount of time required to solve for the parameters of a model is approximately linear to the number of observations. As a rule of thumb, a model with a million observations will take 100 times longer to solve than a model with only 10,000 observations. Further, the amount of variation observed in parameter estimates in "large datasets"—such as those above 100,000 observations—is typically not large enough to warrant a ten fold increase in solution time for each model explored in the analysis. It is also important to recognize that many of the software packages used to estimate discrete choice models were not designed with the primary intention of being applied to millions of observations, and developing more efficient algorithms to solve for parameter estimates on these types of large datasets is still an open area of research. Finally, when "large" datasets are used, many of the statistics (such as *t*-statistics for individual parameter estimates) will be significant at the 0.05 level, making it more difficult to determine which variables should remain in the model (in the sense that they strongly influence choice behavior and improve forecasting accuracy). For these reasons, estimating on samples is recommended. In addition, for some applications such as no show modeling where the choice frequency of "no shows" is small relative to the show choice frequency, choice-based (versus random) sampling offers additional advantages.

To summarize, the key advantage of using samples is related to the reduction in modeling time and the ability to explore a more comprehensive set of model specifications within a fixed project timeline. That is, the use of samples allows the analyst to more quickly identify which variables are important to the choice process. After a preferred model specification has been determined, the analyst can, if desired, estimate the model using the entire "population" sample, and use these parameter estimates in forecasting applications. This is one example of the benefits of using random or non-random samples in airline studies.

Relationship Between Maximum Likelihood Estimators and Sample Dataset

When discrete choice models are used to represent customer behavior, an appropriate estimator must be used to ensure that parameter estimates are consistent (by definition, a consistent estimator is one which converges in probability to the true values of the parameters). In discrete choice models, the selection of an appropriate estimator (readers from an operations research background can loosely think of this as an appropriate "objective function") depends both on the sampling process and the type of choice model. That is, the same estimator that is used to solve for the parameter estimates of a MNL model based on a random sample of observations cannot be directly applied to solve for the parameter estimates of a MNL or NL model based on a choice-based sample of observations. This section presents three maximum likelihood estimators:

the maximum likelihood (ML) estimator, the exogenous sampling maximum likelihood (ESML) estimator, and the weighted exogenous sampling maximum likelihood (WESML) estimator, and explains on which types of samples they should be used. The derivations of ML, ESML, and WESML estimators are provided in Lerman and Manski (1979).

The ML estimator is appropriate to use with any choice model when the estimation data represent a random sample. The ML log likelihood is given as:

$$LL \ for \ ML = \sum_{n=1}^{N} \sum_{i \in C_n} d_{ni} \ln P\left(i \mid x_n, \boldsymbol{\beta}\right)$$

where $P\ (i\ |x, \boldsymbol{\beta}^*)$ is the probability of selecting alternative i given attributes x and parameters $\boldsymbol{\beta}^*$.

The ESML estimator is appropriate to use with any choice model when data are based on exogenous samples, i.e., samples are defined based on the x attributes or any variables other than the observed/stated choice. The log likelihood function used to estimate $\boldsymbol{\beta}$ for exogenous samples is the same as that for random samples except it is calculated by summing over segments, s, and samples in each segment. Formally,

$$LL \ for \ ESML = \sum_{s=1}^{S} \sum_{n=1}^{N_s} \sum_{i \in C_n} d_{ni} \ln P\left(i \mid x_n, \boldsymbol{\beta}\right)$$

Note the utility function can be defined for both random and exogenous samples so that estimated parameters may be allowed to vary across segments. The motivation for collecting an exogenous sample is to ensure that there is enough data so that $\boldsymbol{\beta}$ can be estimated for different segments where appropriate (e.g., different price sensitivities can be modeled for high and low incomes) or to ensure that population groups that are small are adequately represented in the sample.

The WESML estimator is used with choice-based samples to ensure that all parameter estimates are consistent. The log likelihood for a WESML estimator is equivalent to that for the ESML estimator, except that each observation is weighted by the ratio of the alternative's population share, Q_i, to sample share, H_i, or:

$$LL \ for \ WESML = \sum_{n=1}^{N} \sum_{i \in C_n} d_{ni} \left(\frac{Q_i}{H_i} \right) \ln P\left(i \mid x_n, \boldsymbol{\beta}\right)$$

Although the WESML estimator is easy to use, it is not asymptotically efficient, i.e., its variance-covariance matrix does not asymptotically attain the Cramér-Rao bound (Manski and Lerman 1977). From a theoretical perspective, more efficient estimators such as that based on Cosslett (1981) exist, but are generally complicated to implement (e.g., see Brownstone (2001) for a discussion) and have not been

widely adopted. For these reasons, it is common to use the ESML estimator for the case in which a full set of identifiable constants are included in the binary or MNL model. In this special case, it has been shown that when the ESML estimator is used, all parameter estimates, except for the constants, are consistent. In addition, as explained earlier in this chapter under the discussion of alternative specific constants, there is a convenient adjustment procedure that may be applied after the binary or MNL model is estimated.

From a practical perspective, the use of an inefficient estimator such as the WESML can lead to difficulty in determining which attributes are significant for the choice model under study. This is empirically demonstrated in Table 2.7, which compares two MNL models using ESML and WESML estimators. The models use alternative-specific variables to estimate air travelers' day-of-departure rescheduling decisions, i.e., whether passengers show, no show, early standby, or late standby for flights. A standby is defined as a passenger who voluntarily takes a different itinerary with the carrier of interest. Standbys are further divided into those who wait at an airport hoping to take an earlier flight and the late standby who accepts a later flight as a result of missing his/her scheduled flight. The data are from a major network carrier in the U.S. and is based on booking, ticketing, schedule, operating, and check-in data. Sample choice rates are 34 percent, 24 percent, 17 percent and 25 percent and population rates are 92.7 percent, 6.3 percent, 0.89 percent and 0.13 percent for show, no show, early standby and late standby, respectively.

The coefficients of the two models are very similar with the exception of the ASCs. Note that in this example, the ASCs associated with the early standby alternative are stratified with respect to an alternative characteristic of itinerary duration. (The maximum likelihood estimators have the large sample property of consistency, which implies that the estimators approach the true values and therefore, one another, asymptotically; that is, as the sample size approaches infinity. For realistic samples, the estimators are expected to be approximately equal, as seen in the case in this example.) However, the extreme weights included in the log likelihood function lead to increased standard errors in the WESML estimates, as seen in the substantially lower t-statistics. Further, the standard errors obtained using WESML are biased. Excluding constants, all late standby and the majority of early standby variables are statistically insignificant at the 0.05 level in the WESML model, making it difficult to discern which variables influence these choices. For these reasons, it is desirable to use the ESML estimator. To do this, however, the inconsistency in parameter estimates must be quantified (and subsequently eliminated). For a MNL model, this can be achieved when it has a full set of identifiable constants. In this case, McFadden shows that the ESML yields consistent estimators of all parameters except for the constants; moreover, consistent estimators for the constants can be recovered using population share information and subtracting 1n (H_i / Q_i) from the estimated constants. The proof does not require the choice set to be the same across individuals.

Table 2.7 Empirical comparison of weighted and unweighted estimators

	WESML	ESML
Constants (ref. = show)		
Alternative specific constant for NS	1.02 (6.7)	1.33 (11.8)
ASC for ESB: Duration ≤ 180 mins	-4.21 (7.4)	-0.42 (2.0)
ASC for ESB: 180 < duration ≤ 300 mins	-4.48 (6.9)	-0.55 (2.5)
ASC for ESB: Duration > 300 mins	-4.94 (6.5)	-0.87 (3.5)
Alternative specific constant for LSB	-6.63 (6.8)	-0.37 (3.4)
Day of Week (ref. Sun-Tues) Wed–Fri ESB	**0.36 (1.1)**	0.26 (2.9)
Departure Time (ref. = after 7 pm and for NS 6-9 am)		
Depart 9 am – 7 pm NS	-0.28 (2.2)	-0.29 (3.3)
Depart 6 am – 9 am ESB	-1.35 (2.1)	-1.46 (7.2)
Depart 9 am – 4 pm ESB	-1.25 (2.8)	-1.16 (7.3)
Depart 4 pm – 7 pm ESB	**-0.62 (1.3)**	-0.58 (3.6)
Carrier Capacity (100's seats) before scheduled departure for ESB or LSB		
Arrive 1-90 mins earlier	0.25 (2.2)	0.39 (7.0)
Arrive 91-150 mins earlier	**0.30 (1.8)**	0.27 (4.6)
Arrive 151-300 mins earlier	**0.15 (1.4)**	0.07 (2.0)
Arrive 1-90 mins later	**0.11 (0.2)**	**0.08 (1.6)**
Arrive 91-150 mins later	**0.16 (0.4)**	0.15 (3.0)
Arrive 151-300 mins later	**0.11 (0.4)**	0.11 (3.6)
Schedule Presence (Ratio of total flights for carrier vs. nearest competitor)^0.5*		
Departure city presence ESB	**0.12 (0.9)**	0.14 (3.1)
Departure city presence LSB	**0.33 (0.9)**	0.33 (8.8)
E-ticket NS	-1.78 (13.7)	-1.81 (20.2)
Booking Class (ref. = low yield)		
First and business NS	-0.57 (2.0)	-0.74 (3.9)
First and business ESB	**-0.76 (1.1)**	-1.01 (4.2)
First and business LSB	**-0.95 (0.5)**	-1.03 (6.4)
High yield NS	**0.06 (0.4)**	**0.08 (0.7)**

Table 2.7 *Concluded*

	WESML	ESML
High yield ESB	**-0.05 (0.1)**	-0.25 (2.1)
High yield LSB	**-0.35 (0.4)**	-0.46 ((16))
Frequent Flyer (ref. = not a member)		
General member NS	-0.69 (4.1)	-0.54 (4.6)
General member ESB	**0.27 (0.7)**	0.33 (2.4)
General member LSB	**-0.63 (0.7)**	-0.63 (5.6)
Elite member NS	**-0.20 (1.1)**	**-0.13 (1.0)**
Elite member ESB	**0.50 (1.2)**	0.57 (4.2)
Elite member LSB	**-0.49 (0.5)**	-0.47 (3.7)
Group Size (ref. = travel alone)		
Groups of 2-10 individuals NS	-0.45 (3.1)	-0.60 (5.2)
Group of 2-10 individuals ESB	**-0.74 (1.7)**	-0.71 (6.0)
Group of 2-10 individuals LSB	**-0.44 (0.5)**	-0.50 (4.6)
Model Fit Statistics		
LL Zero/LL Constants/LL Model	-4980/-1193/-1067	-3575/-3539/-3112
Rho-Square Zero/Rho-Square Constant	0.786/0.106	0.129/0.121

Key: *S*H=show; ESB= early standby; LSB=late standby.

Notes: Data for March 2001 outbound itineraries for 2,761 observations. Table contains parameter estimate (t-stat). Bold cells are not significant at the 0.05 level. *Applies when carrier of interest has dominant market share.

Source: Adapted from Garrow 2004: Table A1.1 (reproduced with permission of author).

Software Packages for Discrete Choice Estimation

Multiple software packages, such as ALOGIT (ALOGIT Software & Analysis Ltd 2008), BIOGEME (Bierlaire 2003, 2008), ELM (Elm-Works Inc. 2008), Gauss (Aptech Systems Inc. 2008), LIMDEP (Econometric Software Inc. 2008), R (R Development Core Team 2008), SAS (SAS 2008), STATA (StataCorp 2008), and other packages can be used to estimate choice models. Packages including SAS, STATA, and LIMDEP are general econometric packages that include modules for the estimation of MNL in addition to other types of models (such as linear regression and/or time series models).

Packages including ALOGIT, ELM, and BIOGEME were developed with the primary purpose of supporting the estimation of discrete choice models. In contrast, Gauss is a general purpose programming language designed to operate with and on matrices and requires analysts to write their own log likelihood functions.

From a practical perspective, it is important to recognize that different software packages use different input data formats. In general, these software packages differ in terms of whether data should be specified in the idcase-idalt format, or the idcase format. They can also differ in terms of how generic and alternative-specific variables are recognized by (and included in) the software. The difference between the idcase-idalt and the idalt data formats is shown in Tables 2.8 and 2.9, respectively. In the idcase-idalt format, each row contains information about a single alternative, whereas in the idcase format, each row contains information about all alternatives associated with an observation. In the idcase-idalt format, the non-availability of the bus alternative in the first observation is seen by the fact a row for the bus alternative is not included. In the idcase format, the non-availability of the bus alternative is represented by the presence of zeros for all generic (time and cost) variables associated with bus.

The example shown above is meant to illustrate key differences among the two data formats. However, it should be noted that specific requirements differ across software packages. A detailed discussion of the many subtle differences across

Table 2.8 Data in Idcase-Idalt format

OBS (IDCASE)	ALT	d_{ni} (Chosen)	ASC train	ASC bus	Cost ($)	Time (hr)	Male train	Male bus
1	1 Car	0	0	0	$7.00	1.25	0	0
1	2 Train	1	1	0	$5.50	0.33	0	0
2	1 Car	1	0	0	$4.00	0.75	0	0
2	2 Train	0	1	0	$3.00	0.50	1	0
2	3 Bus	0	0	1	$1.75	1.00	0	1
3	1 Car	0	0	0	$5.00	1.00	0	0
3	2 Train	0	1	0	$6.00	0.50	1	0
3	3 Bus	1	0	1	$3.00	0.33	0	1

Table 2.9 Data in Idcase format

OBS (IDCASE)	ALT chosen	Male	Car cost ($)	Car time (hr)	Train cost ($)	Train time (hr)	Bus cost ($)	Bus time (hr)
1	2	0	$7.00	1.25	$5.50	0.33	0	0
2	1	1	$4.00	0.75	$3.00	0.50	$1.75	1.00
3	3	1	$5.00	1.00	$6.00	0.50	$3.00	0.33

different software packages is beyond the scope of this text. However, just like the earlier discussion related to the scale parameter (which may be defined with respect to the "variance" or "inverse variance" of the model), small details related to input data formats are important to keep in mind. All of this highlights the importance of mastering the fundamental concepts in this chapter to avoid errors with incorrectly applying and specifying discrete choice models.

Summary of Main Concepts

This chapter presented fundamental concepts of choice theory and motivated the exploration of more advanced topics contained in the subsequent chapters. The most important concepts covered in this chapter include the following:

- The four fundamental elements defining a choice scenario: the decision-maker, the alternatives available to the decision-maker, attributes of these alternatives, and the decision rule.
- The motivation for using maximum utility theory as a decision rule to represent how individuals make trade-offs among attributes.
- The representation of the utility function as the combination of observed and unobserved components. Different choice models are derived via assumptions on the error terms and/or via assumptions on the distributions of $\boldsymbol{\beta}$'s.
- The underlying sigmoid or S-shape relationship between observed utility and choice probabilities. The S-shape implies that an improvement in the utility associated with alternative i will have the largest impact on choice probabilities when there is an equal probability that alternatives i and j will be selected.
- Fundamental properties of the binary logit and MNL model. The fact that only differences in utility are uniquely identified influences how variables that do not vary across the choice set should be included in the utility function and imposes the need for normalization requirements on error assumptions.
- Adding a constant to utility does not affect which alternative has the maximum utility and does not change choice probabilities. However, although multiplying utility by a constant does not affect which alternative has the maximum utility, it does change the relationship between observed utility and choice probabilities. Choice probabilities are influenced by the amount of variance (represented by the inverse scale) in the model; the higher the variance, the less certain choice probabilities.
- Measures used to interpret the parameters of choice models and/or understand their substitution patterns include odds ratios, derivatives, elasticities, and cross-elasticities.

- The selection of a consistent estimator that will provide unbiased parameter estimates depends both on the type of sample (random, exogenous, choice-based) and the type of model (MNL, NL, etc.).
- Alternative specific constants represent the average effect of factors left out of the model on choice probabilities. When using choice-based samples for binary or MNL models with a full set of ASCs, it is convenient to use a simple (and efficient) unweighted estimator and apply an adjustment to the constants. Similarly, in forecasting applications, it is possible to apply the same adjustment to the constants to match the shares observed in the validation sample.
- Two properties related to the distribution of the maximum of independent Gumbel random variables with the same scale and the distribution of the difference of two independent Gumbel random variables with the same scale will be revisited in the next chapter in the context of NL models.
- One of the main limitations of the MNL model is the "independence of irrelevant alternatives" or IIA property which states that the ratio of choice probabilities between any two alternatives is independent of the availability or attributes of the other alternatives.

The next two chapters focus on relaxations of the IIA property associated with the MNL model for the Generalized Extreme Value (GEV) class of models. GEV models impose the assumption that the total error associated with an alternative follows a Gumbel(0,1) distribution. This enables the choice probabilities to be expressed in closed-form. Mixed logit models, discussed in Chapter 6, also relax the IIA property of the MNL, but require simulation in order to evaluate choice probabilities.

Appendix 2.1: Derivation of MNL Model

This section derives choice probabilities for the multinomial logit (which collapses to the binary logit when the universal choice set has two alternatives). Given the following notation:

n	Individual (observation) index,
C_n	The set of all alternatives for the n^{th} individual,
V_i	Deterministic utility for the i^{th} alternative,
U_i	Total utility for the i^{th} alternative,
ε_i	Error associated with the i^{th} alternative.

The choice probabilities for the MNL model are derived as follows:

$$P_i = P\left(\varepsilon_j < V_i - V_j + \varepsilon_i\right), \quad \forall j \neq i$$

$$P_i = P\left(\varepsilon_1 < V_i - V_1 + \varepsilon_i, \varepsilon_2 < V_i - V_2 + \varepsilon_i, \ldots, \varepsilon_J < V_i - V_J + \varepsilon_i\right)$$

$$P_i = \int_{-\infty}^{\infty} P\left(\varepsilon_1 < V_i - V_1 + \varepsilon_i \mid \varepsilon_i\right) \cdot P\left(\varepsilon_2 < V_i - V_2 + \varepsilon_i \mid \varepsilon_i\right) \cdot \ldots \cdot P\left(\varepsilon_J < V_i - V_J + \varepsilon_i \mid \varepsilon_i\right)$$
$$\cdot f\left(\varepsilon_i\right) \cdot d\varepsilon_i.$$

Using the fact that error terms are independent:

$$P_i = \int_{-\infty}^{\infty} \left[F\left(V_i - V_1 + \varepsilon_i\right) \cdot F\left(V_i - V_2 + \varepsilon_i\right) \cdot \ldots \cdot F\left(V_i - V_J + \varepsilon_i\right)\right] \cdot f\left(\varepsilon_i\right) \cdot d\varepsilon_i$$

$$P_i = \int_{-\infty}^{\infty} f\left(\varepsilon_i\right) \cdot \prod_{\substack{j \in C_n \\ j \neq i}} F\left(V_i - V_j + \varepsilon_i\right) \cdot d\varepsilon_i$$

Imposing the assumption that $\varepsilon_i \overset{\text{IID}}{\sim} G\left(0,1\right)$:

$$P_i = \int_{-\infty}^{\infty} e^{-\varepsilon_i} \cdot e^{-e^{-\varepsilon_i}} \cdot \prod_{\substack{j \in C_n \\ j \neq i}} e^{-e^{-\left(V_i - V_j + \varepsilon_i\right)}} \cdot d\varepsilon_i$$

$$P_i = \int_{-\infty}^{\infty} e^{-\varepsilon_i} \cdot e^{-\left(e^{-\varepsilon_i}\right)} \cdot e^{-\left(\sum_{j \in C_n, j \neq i} e^{-\left(V_i - V_j + \varepsilon_i\right)}\right)} \cdot d\varepsilon_i$$

Define a new variable, w_{i} and introducing new notation for S:

$$w_{\mathrm{i}} = e^{\left[\left(-e^{-\varepsilon_i}\right)\sum_{j\in C_n, j\neq i} e^{-\left(V_i - V_j + \varepsilon_i\right)}\right]} \quad \rightarrow \quad w_{\mathrm{i}} \in [0,1]$$

$$S = \sum_{j\in C_n} e^{-\left(V_i - V_j\right)}$$

Using w_{i} and S, the following relationships hold:

$$w_{\mathrm{i}} = e^{\left(-e^{-\varepsilon_i}\right)\cdot S}$$

$$\frac{dw_{\mathrm{i}}}{d\varepsilon_i} = e^{\left(-e^{-\varepsilon_i}\right)\cdot S} \cdot \frac{d}{d\varepsilon_i}\left[\left(-e^{-\varepsilon_i}\right)\cdot S\right] = e^{\left(-e^{-\varepsilon_i}\right)\cdot S} \cdot \left[e^{-\varepsilon_i}\cdot S\right] = w_{\mathrm{i}} \cdot \left[e^{-\varepsilon_i}\cdot S\right]$$

$$d\varepsilon_i = \frac{1}{w_{\mathrm{i}}\cdot e^{-\varepsilon_i}\cdot S}\cdot dw_{\mathrm{i}}$$

Substitution these relationships back into the probability equation:

$$P_i = \int_0^1 e^{-\varepsilon_i}\cdot w_{\mathrm{i}} \cdot \frac{1}{w_{\mathrm{i}}\cdot e^{-\varepsilon_i}\cdot S}\cdot dw_{\mathrm{i}} = \int_0^1 \frac{1}{\sum_{j\in C_n} e^{-\left(V_i - V_j\right)}}\cdot dw_{\mathrm{i}} = \frac{1}{\sum_{j\in C_n} e^{-\left(V_i - V_j\right)}}$$

Which gives:

$$P_i = \frac{e^{V_i}}{\sum_{j\in C_n} e^{V_j}}$$

Appendix 2.2: Derivation of MNL Elasticities

This section derives elasticities for the MNL model measured at a specific point, X_{ik}. Given the following notation:

X_{ik} The k^{th} attribute w.r.t. the i^{th} alternative,
P_i The probability of the i^{th} alternative,
V_i The deterministic utility of the i^{th} alternative,
β_k The estimated coefficient of the k^{th} attribute,
$\eta_{X_{ik}}^{P_i}$ The elasticity of the probability of the i^{th} alternative w.r.t. the k^{th} attribute, and the following definition for elasticity:

$$\eta_{X_{ik}}^{P_j} = \frac{\partial P_j}{\partial X_{ik}} \cdot \frac{X_{ik}}{P_j}, \quad where \quad P_j = \frac{e^{V_j}}{\sum_{n=1}^{N} e^{V_n}}, \quad and \quad V_j = \sum \beta_k \cdot X_{jk}$$

The derivative of P_i with respect to a change in X_{ik} is:

$$\frac{\partial P_i}{\partial X_{ik}} = \frac{\partial \left[\dfrac{e^{V_i}}{\sum_{j=1}^{J} e^{V_j}} \right]}{\partial X_{ik}} = \frac{\left(\dfrac{\partial e^{V_i}}{\partial X_{ik}} \right) \cdot \left(\sum_{j=1}^{J} e^{V_j} \right) - \left(\dfrac{\partial \sum_{j=1}^{J} e^{V_j}}{\partial X_{ik}} \right) \cdot e^{V_i}}{\left(\sum_{j=1}^{J} e^{V_j} \right)^2} = \frac{e^{V_i} \cdot \beta_k \cdot \left(\sum_{j=1}^{J} e^{V_j} \right) - e^{V_i} \cdot \beta_k \cdot e^{V_i}}{\left(\sum_{j=1}^{J} e^{V_j} \right)^2}$$

$$\frac{\partial P_i}{\partial X_{ik}} = \frac{e^{V_i} \cdot \beta_k}{\left(\sum_{j=1}^{J} e^{V_j} \right)} - \frac{\left(e^{V_i} \right)^2 \cdot \beta_k}{\left(\sum_{j=1}^{J} e^{V_j} \right)^2} = P_i \cdot \beta_k - P^2{}_i \cdot \beta_k = P_i \cdot (1 - P_i) \cdot \beta_k$$

Substituting the derivative into the definition of elasticity gives:

$$\eta_{X_{ik}}^{P_i} = P_i \cdot (1 - P_i) \cdot \beta_k \cdot \frac{X_{ik}}{P_i} = (1 - P_i) \cdot X_{ik} \cdot \beta_k$$

Appendix 2.3: Derivation of MNL Cross-elasticities

This section derives cross-elasticities for the MNL model measured at a specific point, X_{ik}. Given the following notation:

X_{ik} The k^{th} attribute w.r.t. the i^{th} alternative,
P_i The probability of the i^{th} alternative,
V_i The deterministic utility of the i^{th} alternative,
β_k The estimated coefficient of the k^{th} attribute,
$\eta_{X_{ik}}^{P_i}$ The cross-elasticity of the probability of the j_{th} alternative w.r.t. the k^{th} attribute of the i^{th} alternative, and the definition of cross-elasticity:

$$\eta_{X_{ik}}^{P_j} = \frac{\partial P_j}{\partial X_{ik}} \cdot \frac{X_{ik}}{P_j}, \quad where \quad P_j = \frac{e^{V_j}}{\sum_{n=1}^{N} e^{V_n}}, \quad and \quad V_j = \sum \beta_k \cdot X_{jk}$$

The derivative of P_j with respect to a change in X_{ik} is:

$$\frac{\partial P_j}{\partial X_{ik}} = \frac{\partial\left[\frac{e^{V_j}}{\sum_{n=1}^{N} e^{V_n}}\right]}{\partial X_{ik}} = \frac{\left(\frac{\partial e^{V_j}}{\partial X_{ik}}\right)\cdot\left(\sum_{n=1}^{N} e^{V_n}\right) - \left(\frac{\partial \sum_{n=1}^{N} e^{V_n}}{\partial X_{ik}}\right)\cdot e^{V_j}}{\left(\sum_{n=1}^{N} e^{V_n}\right)^2} = \frac{0\cdot\left(\sum_{n=1}^{N} e^{V_n}\right) - e^{V_i}\cdot\beta_k\cdot e^{V_j}}{\left(\sum_{n=1}^{N} e^{V_n}\right)^2}$$

$$\frac{\partial P_j}{\partial X_{ik}} = -\frac{e^{V_i}}{\sum_{n=1}^{N} e^{V_n}} \cdot \frac{e^{V_j}}{\sum_{n=1}^{N} e^{V_n}} \cdot \beta_k = -P_i \cdot P_j \cdot \beta_k$$

Substituting the derivative into the definition of elasticity gives:

$$\eta_{X_{ik}}^{P_j} = -P_i \cdot P_j \cdot \beta_k \cdot \frac{X_{ik}}{P_j} = -P_i \cdot \beta_k \cdot X_{ik}$$

Note that the IIA property is illustrated by the fact that the cross-elasticity does not depend on P_j.

This page has been left blank intentionally

Chapter 3
Nested Logit Model

Introduction

Among all discrete choice models, the MNL is the model that is most frequently used in practice. The key strengths of the MNL include the simplicity of its probability expression and the ability to leverage the IIA property to forecast share shifts due to the addition or removal of alternatives from the choice set. Both of these strengths arise from the assumption that error terms are independently and identically distributed (iid) and follow a Gumbel distribution. However, as discussed in the previous chapter, these same assumptions also impose important behavioral limitations. In particular, the MNL model cannot incorporate random taste variation and is not appropriate to use in situations in which error terms are correlated across observations, as is the case with panel data or data that contain multiple responses from the same individual. In addition, the IIA property, although particularly useful for forecasting share changes due to the addition or removal of alternatives, can result in unrealistic substitution patterns across alternatives. The nested logit (NL) model, which appeared just a few years after the MNL model (Williams 1977; McFadden 1978), incorporates more realistic substitution patterns by relaxing the assumption that error terms are independent. Consistent with the MNL model (and all models belonging to the Generalized Extreme Value class), the NL model maintains the assumption that the total error associated with an alternative is identically distributed; this assumption of equal variance across alternatives is required in order to obtain a closed-form probability expression. Several other assumptions inherent to the MNL also appear in the context of the NL model, i.e., the NL model assumes that there is no random taste variation and that errors are independent across observations. Despite these limitations, however, the NL model is the second most frequently used choice model in practice. Within the airline industry, there are many applications in which the NL model can offer forecasting benefits over the MNL model. These include the no show model presented in the previous chapter and itinerary choice models (covered in Chapter 7) in which the NL model is used to incorporate increased substitution among itineraries that belong to the same carrier, departure time, and/or level of service.

This chapter presents fundamental concepts related to the NL model. The next section provides interpretations for NL probability expressions, correlations, elasticities, and cross-elasticities. Next, two in-depth examples are covered. The first example, based on airline passengers' willingness to pay, is used to reinforce the interpretation of NL probabilities and correlations. The second

example focuses on methods used to generate synthetic NL datasets, an issue that is of general concern for the travel demand modeling community. It also serves to further highlight subtle interpretations that can arise from underlying assumptions related to error components. Armed with a solid understanding of how assumptions on error components relate to the definition of the NL model, a cautionary note is provided on the use of different kinds of "nested logit" models mentioned in the literature and used in practice. The chapter concludes with two technical appendices. The first appendix derives NL probabilities and the second derives NL correlation.

NL Choice Probabilities

Similar to the MNL, the NL is a choice model that is used to predict the probability that an individual will select one alternative out of a set of mutually exclusive and collectively exhaustive alternatives. Both MNL and NL models are based on random utility theory, but differ in how they represent substitution patterns among alternatives. This is accomplished via different assumptions related to error terms. As discussed in Chapter 2, the utility function for alternative i and individual n is expressed in the MNL model as:

$$U_{ni} = V_{ni} + \varepsilon_{ni}$$

where U_{ni} is the true utility (unknown to the analyst) expressed as the sum of an observed component, V_{ni}, and an unobserved component, ε_{ni}. MNL probabilities are derived by assuming $\varepsilon \sim$ iid $G(0,\gamma)$ across alternatives and individuals. The iid assumptions can be clearly observed in the variance-covariance matrix, Ω. Given a choice set with four alternatives, the associated variance-covariance matrix is expressed as:

$$\Omega = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{bmatrix} \frac{\pi^2}{6\gamma^2} & 0 & 0 & 0 \\ 0 & \frac{\pi^2}{6\gamma^2} & 0 & 0 \\ 0 & 0 & \frac{\pi^2}{6\gamma^2} & 0 \\ 0 & 0 & 0 & \frac{\pi^2}{6\gamma^2} \end{bmatrix} = \frac{\pi^2}{6\gamma^2} \times \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The assumption that errors are distributed independently across alternatives is seen by the fact that all off-diagonal covariance terms are zero. The assumption

that errors are identically distributed is seen by the fact that all diagonal terms have the same variance, or $\pi^2/6\gamma^2$.

The NL model relaxes the assumption that errors are independently distributed by grouping alternatives into M nests, i.e., $i \in A_m$, $m = 1,2,\ldots, M$. An alternative belongs to one and only one nest. The NL utility function can be expressed as follows (suppressing the index for individual n for notational convenience):

$$U_{im} = V_i + \varepsilon_m + \varepsilon_i$$

That is, the total variance associated with each alternative in nest m is decomposed into a common error component, ε_m , and an independent error term, ε_i. Alternatives that belong to the same nest share a common error term, whereas alternatives that are in different nests have independent error terms. Conceptually, there are two assumptions associated with the distribution of error terms for the NL model. The first assumption states that the total variance associated with each alternative, given as the sum of ε_m and ε_i must be identically distributed and follow a Gumbel with mode zero and scale γ, implying a total variance of $\pi^2/6\gamma^2$. The second assumption states that the independent error terms, ε_i, also follow a Gumbel with mode zero, but with a different scale. Formally, ε_i are distributed such that they have a cumulative distribution function of

$$\exp\left(-\sum_{m=1}^{M}\left(\sum_{i\in A_m}\exp^{-\varepsilon_i/\mu_m}\right)^{\mu_m}\right), 0<\mu_m\le 1$$

The variance associated with the independent error component is $\pi^2\mu_m^2/6\gamma^2$. The logsum parameter, μ_m, is a measure of the degree of correlation and substitution among alternatives in nest m. Higher values of μ_m imply less, and lower values imply more, correlation among alternatives in the nest. In turn, higher correlation leads to greater competition effects among alternatives in the nest.

The variance of the common error component is given as the difference between the total variance and the independent variance and is derived by noting that the "common" and "independent" error terms are independent. Formally, the variance of the common error component is obtained using the relationship for the variance of a sum of independent random variables, or

$$\text{Var(common)} = \text{Var(total)} - \text{Var(independent)} = \frac{\pi^2}{6\gamma^2}\left(1-\mu_m^2\right)$$

It is important to note that although the *variance* of the common error component can be easily derived, the *distribution* associated with the common error component is not as clear. As described in Chapter 2, the distribution associated with the common error component is given by the difference of two

Gumbel distributions *with different scales* and falls “somewhere” between the Gumbel and logistic distributions (shown in Figures 2.6 and 2.8). This point will be revisited when discussing methods that can be used to generate synthetic NL datasets for simulation experiments.

An example of a NL model with four alternatives and two nests is shown in Figure 3.1. Normalizing the scale of γ to one for notational convenience, the variance-covariance matrix associated with this NL model is given as:

$$\Omega = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} \frac{\pi^2}{6} & \frac{\pi^2}{6}\left(1-\mu_1^2\right) & 0 & 0 \\ \frac{\pi^2}{6}\left(1-\mu_1^2\right) & \frac{\pi^2}{6} & 0 & 0 \\ 0 & 0 & \frac{\pi^2}{6} & \frac{\pi^2}{6}\left(1-\mu_2^2\right) \\ 0 & 0 & \frac{\pi^2}{6}\left(1-\mu_2^2\right) & \frac{\pi^2}{6} \end{bmatrix} \end{array}$$

From the variance-covariance matrix, it is clear that the total variance of all alternatives is the same (since the diagonal elements are identical). Alternatives that belong to the same nest share a common error term (i.e., have a covariance term), whereas alternatives that are in different nests have independent error terms (i.e., their covariance is zero).

Under the assumptions that $(\varepsilon_i + \varepsilon_m)$ is identically distributed $G\,(0,\gamma)$ and that ε_i is distributed $G\,(0,\gamma / \mu_m)$ $i \in A_m$, $m = 1,2,\ldots, M$, the probability that individual n selects alternative i is given as:

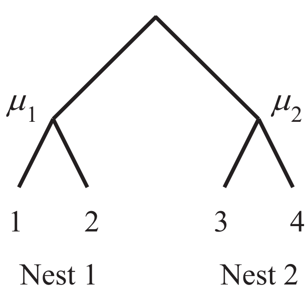


Figure 3.1 Example of a NL model with four alternatives and two nests

$$P_i = \frac{e^{V_i/\mu_m}\left(\sum_{j\in A_m} e^{V_j/\mu_m}\right)^{\mu_m - 1}}{\sum_{l=1}^{M}\left(\sum_{j\in A_l} e^{V_j/\mu_l}\right)^{\mu_l}}, \quad 0 < \mu_m \le 1 \tag{3.1}$$

A more intuitive expression for the NL choice probability can be derived as the product of a conditional and marginal probability (this derivation is provided in Train, 2003, p. 90). This formulation is particularly helpful when extending NL models to include additional levels of nests.

$$P_i = P_{i|m} \times P_m = \frac{e^{\left(\frac{V_i}{\mu_m}\right)}}{\sum_{j\in A_m} e^{\left(\frac{V_j}{\mu_m}\right)}} \times \frac{e^{V_m + \mu_m \Gamma_m}}{\sum_{l=1}^{M} e^{V_l + \mu_l \Gamma_l}}, \quad \Gamma_m = \ln\left(\sum_{j\in A_m} e^{\left(\frac{V_j}{\mu_m}\right)}\right), \quad 0 < \mu_m \le 1$$

The first component of the product is the probability of selecting alternative i among all j alternatives in nest m, conditional on the choice of m, and the second product is the probability of selecting nest m among all nests. Γ_m is often called the "log-sum term" because it is the log of a sum (this terminology should not to be confused with μ_m, the "logsum" or "logsum parameter"). The log-sum term is frequently used in urban travel demand applications to provide linkages between model components (e.g., the log-sum term from a mode choice model may be incorporated into a destination choice model). However, to the author's knowledge, there are no aviation applications that use the "log-sum term" to link model components.

Interpretation of Correlation

In NL models, the logsum parameter, μ_m, plays a very important role in interpretation. As previously mentioned, μ_m is a measure of the degree of correlation and substitution among alternatives in nest m. Formally, correlation among the alternatives in a common nest is given as the ratio of the common variance to the total variance, or $\rho^2 = \left(1 - \mu_m^2\right)$. Given that logsum parameters are bounded from zero to one, it is clear from the formula that values close to one imply less, and values close to zero imply more, correlation among alternatives in the nest. Note that a value of $\mu_m = 1$ for all nests is equivalent to a MNL model (no correlation across alternatives). This relationship will be used in Chapter 7 to develop statistical tests that can be used to compare the fits of MNL and NL models.

The requirement that logsum parameters are bounded between zero and one is motivated by two theoretical properties. First, logsum values outside of the (-1,1) range are not theoretically valid as they would result in a negative variance for the independent or common error components.[1] Second, the range of (0,1] is required to theoretically ensure that the NL model is consistent with utility maximization. Conceptually, the NL model groups alternatives that the analyst hypothesizes share common, unobserved attributes (or, stated another way, that have a *positive covariance*). These unobserved attributes cannot be incorporated into the observed portion of utility. A classic example from the mode choice literature is to place public transit modes (such as bus and train) in the same nest to capture common characteristics that are difficult to measure and forecast—such as lack of privacy and unreliability in schedules. In the no show model discussed in the context of WESML estimators in Chapter 2, the alternatives associated with the show, early standby, and late standby choices were all placed in the same nest to capture common unobserved factors related to the fact that, unlike no show customers, show and standby customers all went to the airport with the intention of traveling on the carrier of interest.

From an interpretation perspective, the incorporation of positive covariance among alternatives that share a common nest also leads to increased substitution among these alternatives. In forecasting applications, this means that an improvement in an alternative will draw proportionately more share from alternatives that belong to the same nest than from alternatives that belong to different nests. As an example, consider an itinerary choice application in which nests are created that group itineraries by operating carrier. If Delta improves one of its itineraries (e.g., by changing the schedule so that it operates at a time that is more desirable to passengers), the NL model will predict that the increased share associated with this itinerary is due to drawing proportionately more passengers from existing Delta flights than from those of its competitors. Elasticity and cross-elasticity formulas, discussed in the next section, are used to formally express this increased substitution as a function of the logsum parameter(s). Conceptually, increased substitution among alternatives in the nest only occurs when they have a positive covariance, which requires that the logsum parameter is between zero and one. In the case of negative covariance, an improvement in one alternative may result in a decrease in share for that alternative, which is not consistent with utility maximization.

Interpretation of Elasticities and Cross-elasticities

Direct- and cross-elasticities are used to examine and understand the substitution patterns of MNL and more complex discrete choice models. Table 3.1 compares the direct- and cross-elasticities associated with a percentage

1 As an aside, theoretically it has been shown that, under certain conditions, logsum parameters can be larger than one (e.g., see Kling and Herriges 1995; Herriges and Kling 1996; Train 1987). However, in practical applications, these logsum parameters are always restricted to the (0,1] range.

change in the k^{th} attribute for alternative i. The direct-elasticity measures the "direct effect" on P_i associated with a change in a variable in the utility function for alternative i. Similarly, the cross-elasticities measures the "indirect" or "cross" effect on P_j associated with a change in a variable in the utility function for alternative i.

There are several points of interest in Table 3.1. First, when $\mu_m = 1$, the NL direct- and cross-elasticities are identical to those of the MNL model. Second, even when $0 < \mu_m < 1$, the MNL and NL cross-elasticities are identical for alternatives that are not in the same nest. Intuitively, this is because alternatives that are not in the same nest do not share a common error term and are independent. Thus, for alternatives in different nests, the independence of irrelevant alternatives (IIA) property associated with the MNL model holds. The term "independence of irrelevant nests (IIN)," coined by Train (2003), is useful when describing this property. Third, when $0 < \mu_m < 1$, the NL direct- and cross-elasticities show that alternatives in nest m are more sensitive to changes in X_{ik} than with the MNL model. Specifically, an improvement in alternative i that belongs to nest m will draw proportionately more share from other alternatives in nest m than from alternative that are not in nest m. It is important to note that the forecasts from a NL model will not necessarily result in "more share" for alternative i when compared to forecasts from a MNL model. That is, the NL model simply states that an improvement in alternative i will draw *proportionately* more customers from alternatives in the same nest. In the case when the shares associated with the alternatives in the nest are small (i.e., there is a "smaller pool" of customers that are realistically expected to change their choices), the NL model effectively helps protect against over-forecasting the effect of improving alternative i.

Table 3.1 Comparison of direct- and cross-elasticities for MNL and NL models

	MNL	NL
Direct	$\eta_{X_{ik}}^{P_i} = (1-P_i)\beta_k X_{ik}$	$\eta_{X_{ik}}^{P_i} = \left[(1-P_i) + \left(\frac{1-\mu_m}{\mu_m}\right)(1-P_{i\|m})\right]\beta_k X_{ik}$
Cross	$\eta_{X_{ik}}^{P_j} = -P_i \beta_k X_{ik}$	$\eta_{X_{ik}}^{P_j} = -P_i \beta_k X_{ik}$ for *i, j* in different nests $\eta_{X_{ik}}^{P_j} = -\left[P_i + \left(\frac{1-\mu_m}{\mu_m}\right)P_{i\|m}\right]\beta_k X_{ik}$ for *i, j* in same nest

Extension to Three-level NL Models

The concepts presented in the context of the "standard" NL model shown in Figure 3.1 can be easily extended to NL models with three or more levels to incorporate even more flexible substitution patterns. An example of a three-level NL model is shown in Figure 3.2. Note that in this discussion, *n* refers to a second-level nest, not an individual as used in previous discussions. The probability of choosing alternative *i* in a three-level nest is given as:

$$P_i = P_{i|nm} \times P_{nm|m} \times P_m = \frac{e^{V_i/\mu_{nm}}}{\sum_{j\in B_n} e^{V_j/\mu_{nm}}} \times \frac{e^{\mu_m \times \mu_{nm}\Gamma_{nm}}}{\sum_{k\in A_m} e^{\mu_m \times \mu_{km}\Gamma_{km}}} \times \frac{e^{\mu_m \Gamma_m}}{\sum_{l=1}^{M} e^{\mu_l \Gamma_l}},$$

$$\Gamma_{nm} = \ln\left(\sum_{j\in B_n} e^{V_j/\mu_{nm}}\right), \quad \Gamma_m = \ln\left(\sum_{k\in A_m} e^{\mu_m \times \mu_{km}\Gamma_{km}}\right), \quad 0 \le \mu_m, \mu_{nm} \le 1$$

The first component of the product is the probability of selecting alternative *i* among all *j* alternatives in nest *nm*, conditional on the choice of *nm*. The second product is the probability of selecting nest *nm* among all two-level nests in nest *m*, conditional on the choice of *m*. The third product is the probability of selecting nest *m* among all three-level nests.

In order to ensure that the three-level NL model is consistent with utility maximization, the correlation must increase as one moves down the tree. In Figure 3.2, this implies that the correlation (or substitution among alternatives) is greater between alternatives 2 and 3 than it is between alternatives 2 and 5. Formally, $\mu_{12} \le \mu_2$, $\mu_{13} \le \mu_3$ and $\mu_{23} \le \mu_3$. The variance-covariance matrix associated with this model is shown below Figure 3.2 (note that for notational convenience, the common term of $\pi^2/6\gamma^2$ has been factored out and only the upper diagonal is shown). Extension to nests with four or more levels is straightforward, albeit the estimation of these models becomes more involved, as discussed in the next section.

$$\Omega = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 1 & 1-\mu_{12}^2 & 1-\mu_{12}^2 & 1-\mu_2^2 & 0 & 0 & 0 & 0 \\ & & 1 & 1-\mu_{12}^2 & 1-\mu_2^2 & 0 & 0 & 0 & 0 \\ & & & 1 & 1-\mu_2^2 & 0 & 0 & 0 & 0 \\ & & & & 1 & 0 & 0 & 0 & 0 \\ & & & & & 1 & 1-\mu_{13}^2 & 1-\mu_3^2 & 1-\mu_3^2 \\ & & & & & & 1 & 1-\mu_3^2 & 1-\mu_3^2 \\ & & & & & & & 1 & 1-\mu_{23}^2 \\ & & & & & & & & 1 \end{bmatrix} \times \frac{\pi^2}{6\gamma^2}$$

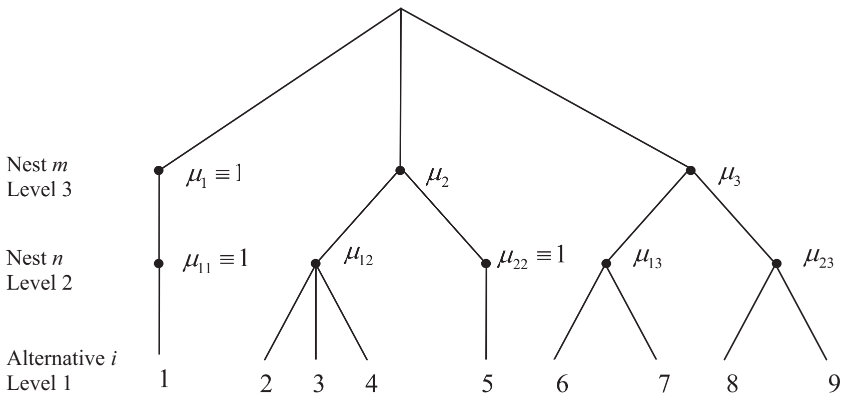


Figure 3.2 Example of a three-level NL model

Estimation Considerations

There are many questions the analyst must answer when estimating NL models: How should alternatives be grouped? How do we ensure logsum parameters are in the required ranges and exhibit appropriate relationships among levels of the nest? How do we estimate models that contain hundreds of alternatives, some of which may be chosen very infrequently in the sample? The problem is further complicated when considering that, unlike the MNL model, the NL log likelihood function is no longer globally concave, and the maximization of the log likelihood function is effectively a non-linear optimization problem. Airline applications pose additional unique challenges, as the data available for estimation are often one or two orders of magnitude larger than that encountered in marketing, economics, and urban travel demand areas, i.e., areas that drove the initial development of optimization algorithms for these models.

In practice, the analyst spends approximately 80 to 90 percent of the modeling effort on estimating MNL models and developing an intuitive, well-specified utility function that explains how behavioral factors influence choices. Given a well-specified utility function, advanced models that incorporate more realistic substitution patterns are then explored. Due to the interaction between the systematic and unobserved components of utility, it is possible that the interpretation of the systematic portion will change when more advanced models are estimated; however, in most cases parameter estimates are fairly stable across MNL and NL model estimations. Differences in MNL and NL parameter estimates primarily arise with alternative-specific variables, particularly those associated with an alternative that is infrequently chosen. Intuitively, this is because when an alternative is chosen infrequently, it is difficult to obtain a stable estimate, i.e., determine how this variable influences the probability the alternative will be chosen.

When estimating NL models, researchers generally use unconstrained optimization methods and verify *ex ante* that logsum coefficients are in the appropriate range and exhibit the appropriate relationships among levels of the nest. In cases where the logsums fall outside the (0,1] range,[2] the starting values can be changed (to see if the optimization algorithm stopped at a local optimum and/or to see if there is a local optimum that satisfies the appropriate constraints). More commonly, however, the nesting structure that resulted in the invalid logsum parameter estimate(s) is rejected from further consideration. In cases where there are numerous alternatives, it is also common to constrain logsum parameters at the same level of the tree to be equal to each other.

Thus, in practice, the estimation of NL models involves a strong mix of analyst judgment combined with "optimization tricks." However, given that the development of non-linear optimization algorithms for these models primarily occurred during the 1980's in the economics community and prior to the availability of large datasets with hundreds, if not thousands of alternatives, the author believes that there are many research opportunities for developing more efficient and robust optimization algorithms (perhaps involving sparse matrix techniques) for logit model estimation. It would also be helpful to develop automated methods to determine nesting structures with the best fit because analysts currently estimate nesting structures one-by-one by "manually" setting up these nests. A detailed example of the modeling process used to develop MNL and NL models for itinerary choice applications is provided in Chapter 7.

Example: Willingness to Pay

Data and Model Formulation

At this point in the discussion, it is helpful to present an empirical example that shows how to use the NL formula to calculate probabilities. The example, adapted from Garrow, Jones, and Parker (2007) and reproduced with permission of Palgrave Macmillan, is based on data from a stated preference survey. Specifically, information about individuals' willingness to pay to travel by air and itinerary choice was obtained via a stated preference survey conducted in 2004 from consumers using an Internet-based airline ticket booking service that searched for fares across multiple sites. While waiting for the search engine to return airline flights and fares, customers were asked to complete a short survey tailored to the outbound origin and destination about which they were seeking itinerary information. Each respondent was shown one choice set and was asked to rank the three alternatives

2 The notation (0,1] is used as logsum coefficients may take on the value of one (represented by the use of the inclusive "]" bracket, but not the value of one (represented by the use of the "(" bracket. A value of one for all logsum coefficients is equivalent to a MNL model.

offered. Respondents also indicated whether they would fly, not take the trip, or take the trip by a different mode (e.g., car, train, bus) if these were the only three air alternatives that were available.

From a modeling perspective, it is desirable to distinguish between the influence of price on the decision to travel (or modeling *how many* people will travel) and the influence of price on itinerary choices (or modeling the share among itineraries). Formally, *willingness to pay to travel* is defined as one's willingness to pay to travel by air (or the additional demand that is stimulated to travel by air when air fares decrease) and *willingness to pay for airline service* is defined as one's willingness to pay for itinerary services (e.g., the fare premium airline passengers are willing to pay for service characteristics such as traveling on a non-stop flight, a preferred airline, etc.). These two components of willingness to pay are modeled using a formulation similar in spirit to one recently used by Subramanian (2006) in the context of joint estimation of destination and mode choice. Specifically, two price variables are defined. The first variable is defined as the average itinerary price in the choice set for individual n, $\overline{p}^n$, and is common to all air itineraries in the choice set. In contrast, the second variable varies across all air itineraries in the choice set and is defined as the difference between the price of itinerary i and the average price in the choice set, or $p_i^n - \overline{p}^n$. The average itinerary price captures the willingness to pay to travel by air. The deviation from the average price captures the willingness to pay for itinerary services. For example, assume three itineraries with fares of \$300, \$250, and \$200. The observed utility associated with the price variables for the three itineraries and the no fly reference, V_4^n, is given as:

$$V_i^n = \beta_1\left(Average\ Price^n\right) + \beta_2\left(Price_i^n - Average\ Price^n\right)$$

$$V_1^n = \beta_1\left(\$250\right) + \beta_2\left(\$300 - \$250\right)$$

$$V_2^n = \beta_1\left(\$250\right) + \beta_2\left(\$250 - \$250\right)$$

$$V_3^n = \beta_1\left(\$250\right) + \beta_2\left(\$200 - \$250\right)$$

$$V_4^n = 0$$

The representation of time follows a formulation similar to that of price, except that the minimum travel time (represented as distance) is used as the reference. Two components are used. The first variable is distance and applies to all air itineraries. Distance is highly, but not perfectly, correlated with the non-stop travel time in the market. This is because non-stop travel time is measured as "gate-to-gate" time, which includes the taxi-out, flight, and taxi-in times, and because in many cases flight paths are not perfectly point-to-point. The second variable is incremental flight time over the non-stop flight and measures the additional flight time due to a connecting itinerary. For example, if three itineraries from San Francisco to Boston with a distance of 2,696 miles have total air trip times of 330, 480, and 420

minutes, respectively, then the observed utility associated with the distance and time variables for the three itineraries and the no fly reference, V_4^n, is given as:

$$V_i^n = \beta_3\left(Distance^n\right) + \beta_4\left(Total\ Air\ Trip\ Time_i - Non\text{-}stop\ Time^n\right)$$

$$V_1^n = \beta_3\left(2{,}696\right)$$

$$V_2^n = \beta_3\left(2{,}696\right) + \beta_4\left(480 - 330\right)$$

$$V_3^n = \beta_3\left(2{,}696\right) + \beta_4\left(420 - 330\right)$$

$$V_4^n = 0$$

NL parameter estimates are shown in Table 3.2. Parameter estimates for departure and arrival time preferences, income, booking date, and length of stay are suppressed from the example, but contained in the Garrow, Jones, and Parker (2007) paper. The logsum coefficient shown in the table reflects the NL model structure shown in Figure 3.3. Intuitively, the non-stop, connection on same airline, and connection on different airline are grouped into the same nest to reflect the hypothesis that passengers who chose an itinerary (or decided to travel by air) are more likely to switch to a different itinerary than decide not to travel by air. The logsum coefficient of 0.27 indicates a high degree of correlation among these alternatives ($1\text{-}0.27^2 = 0.93$), supporting the hypotheses that these alternatives should be grouped in the same nest. Alternative nesting structures are also possible, such as a three-level NL model that imposes the highest amount

Table 3.2 NL model results for willingness to pay

	NL Model
Constants (reference = no air travel)	
ASC[1] non-stop	1.08 (10.2)
ASC connection on same airline	0.75 (5.2)
ASC connection on different airline	0.70 (4.6)
Price (hundreds of dollars)	
Leisure & self-pay business: Average price in choice set *willingness to pay to travel by air*	–0.41 (16.6)
Leisure & self-pay business: Price – avg. price in choice set *willingness to pay for service tradeoffs*	–0.56 (5.7)
Reimbursed business: Average price in choice set *willingness to pay to travel by air*	–0.23 (6.8)

Table 3.2 *Concluded*

	NL Model
Reimbursed business: Price – avg. price in choice set *willingness to pay for service tradeoffs*	–0.29 (4.3)
Distance (hundreds of miles)	
Min air distance (applies to all air options)	0.06 (9.2)
Incremental flight time (hours)	
Leisure and self-pay business: Total air – non-stop flight time	–0.11 (3.3)
Reimbursed business: Total air – non-stop flight time	–0.20 (3.9)
NL logsums	
Mu 1: Air competition nest (see Figure 4.3)[2]	0.27 (15.1)
Model fit statistics	
LL Zero / LL Constants	
LL Model	–2954.72
Rho-Squared$_{zero}$/Rho-Squared$_{constant}$	0.267/0.158
Number of cases / Number of variables	2907/23
Value of time	
Leisure and business self-pay	\$19.64/hr
Reimbursed business trips	\$68.97/hr

Notes: Parameter (t-stat). [1]Alternative specific constant. [2]T-stat reported against 1.

Source: Adapted from Garrow Jones and Parker 2007: Table 2 (reproduced with permission of Palgrave Macmillan).

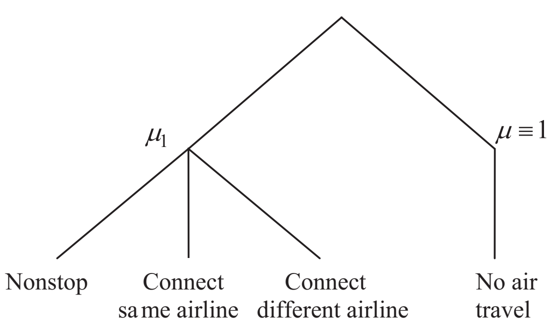


Figure 3.3 NL model of willingness to pay

Source: Garrow Jones and Parker 2007: Figure 2 (reproduced with permission of Palgrave Macmillan).

of competition among connecting itineraries, and greater-than-MNL competition among all air itineraries.

NL Probability Calculation

NL probabilities are calculated for a "representative" passenger, defined as an individual traveling from SFO-BOS with a distance of 2,696 miles. Both connecting itineraries are 60 minutes longer than the non-stop. The observed NL utility for alternative one is:

$$V_1 = \beta_{Non\text{-}Stop} + \beta_1\left(Average\ Price\right) + \beta_2\left(Price_1 - Average\ Price\right)$$
$$+ \beta_3\left(Distance\right) + \beta_4\left(Total\ Air\ Trip\ Time_1 - Non\text{-}stop\ Time\right)$$

Using the coefficients and attributes for the representative leisure passenger described earlier who is choosing among fares of \$300, \$250, and \$200 on the non-stop, connection same airline, connection different airline itineraries, respectively, the observed utilities for each of the alternatives are:

$$V_1 = 1.08 - 0.41\left(2.5\right) - 0.56\left(0.5\right) + 0.06\left(26.96\right) - 0.11\left(0\right)$$
$$= 1.3926$$

$$V_2 = 0.75 - 0.41\left(2.5\right) - 0.56\left(0\right) + 0.06\left(26.96\right) - 0.11\left(1\right)$$
$$= 1.2326$$

$$V_3 = 0.70 - 0.41\left(2.5\right) - 0.56\left(-0.5\right) + 0.06\left(26.96\right) - 0.11\left(1\right)$$
$$= 1.4626$$

$$V_4 = 0$$

Note that a price of \$250 is divided by 100 and entered as 2.5 (hundreds of dollars) in the observed utility function. Similarly, an incremental flight time of 60 minutes is entered as 1.0 (hours) and a distance of 2,696 miles is entered as 26.96 (hundreds of miles). This rescaling is done to help the optimization software more quickly solve for parameter estimates. That is, some—but not all—off-the-shelf optimization procedures embedded in logit modeling software will automatically rescale variables so that all parameter estimates are within the same order of magnitude (which speeds up the optimization process).

The NL probabilities can be calculated using Equation 3.1. Note that the logsum coefficients associated with each nest are defined as:

μ_1= logsum for air nest = 0.27

μ_2= logsum for no fly (degenerate) nest = 1

Using this information, e^{V_i/μ_k} can be calculated for each alternative to give:

$$e^{V_1/\mu_1} = \quad e^{1.3926/0.27} = 173.78$$

$$e^{V_2/\mu_1} = \quad e^{1.2326/0.27} = 96.08$$

$$e^{V_3/\mu_1} = \quad e^{1.4626/0.27} = 225.21$$

$$e^{V_4/\mu_2} = \quad e^{0/1} = 1$$

Finally, the probability for alternative 1 (non-stop) is given as:

$$P_1 = \frac{e^{V_1/\mu_1}\left(e^{V_1/\mu_1} + e^{V_2/\mu_1} + e^{V_3/\mu_1}\right)^{\mu_1 - 1}}{\left(e^{V_1/\mu_1} + e^{V_2/\mu_1} + e^{V_3/\mu_1}\right)^{\mu_1} + \left(e^{V_4/\mu_2}\right)^{\mu_2}} = \frac{173.78 \times (173.87 + 96.08 + 225.21)^{0.27-1}}{(173.87 + 96.08 + 225.21)^{0.27} + (1)^1}$$
$$= 0.296$$

Value of Time Calculation

In order to interpret the parameter estimates associated with time and cost, it is common to calculate value of speed (or equivalently, value of time, as it is more commonly referred to in the transportation literature). Value of time is defined as the amount of money individuals are willing to spend to save one hour of travel time. Under the assumption that the observed utility function is linear in price, the dollar value of time expressed as dollars per hour is calculated using the $\boldsymbol{\beta}$ coefficients associated with the incremental flight time and incremental price variables. Because utility is dimensionless, the units associated with β_2 of $\beta_2\left(Price_i^n - Average\ Price^n\right)$ is 1/(100\$), whereas the units associated with β_4 of β_4 (*Total Air Trip Time*$_i$ – *Non-stop Time*^{n}) is 1/hour. The value of time for the leisure traveler for the NL model shown in Table 3.2 is given as:

$$\frac{\beta_4}{\beta_2} \times 100 = \frac{-0.11}{-0.56} \times 100 = \$19.64$$

Compared to other transportation models, relatively little is known or published about airline passengers' willingness to pay. For example, in 2001 Hensher summarized value of time studies for intercity travel reported in the literature. His review, which synthesized results from more than 60 studies and nine countries, found that value of travel time savings estimates are highly context dependent (i.e., they depend on mode, trip length, trip purpose, etc.) and that the "vast majority of available analysis in the literature are for contexts other than air travel, and the availability of value of travel time savings figures for air travel is limited" (Hensher 2001c). Further, among the limited intercity studies that exist, Hensher found that not all of the evidence was "reinforcing" and emphasized the need to conduct new empirical studies targeted to specific markets in which a new aircraft will operate.

In the Garrow, Parker, and Jones (2007) study used for this example, the leisure value of time for transcontinental flights ($19.64) is lower than the $31 per hour obtained from Hensher's synthesis of the literature which was based primarily on studies conducted prior to 2000 and before the penetration of on-line distribution channels. The value of time has not been adjusted for inflation; Hensher's result assumes U.S. currency from 2000. However, in contrast to values of time for leisure passengers, the value of time for business passengers ($68.97) is slightly higher than the $52 per hour estimate obtained by Hensher. Thus, to summarize, the values of time obtained with the new price formulation and stated preference data from travelers searching for information on-line are comparable to prior studies and provide qualitative evidence that reinforces industry perceptions that: 1) on-line leisure passengers are more price sensitive than other leisure passengers, yet 2) there are still opportunities for premium pricing via on-line channels provided one can determine ways to segment leisure and business travelers.

Example: Generation of Synthetic Datasets for NL Models

This section presents different methods that have been used in the travel demand modeling community to generate synthetic NL datasets. There are two primary motivations for this section. First, the discussion emphasizes subtle interpretation nuances that arise due to distribution assumptions associated with the error components. Second, this section highlights open research questions and opportunities the author believes that airline practitioners and researchers trained in simulation analysis are particularly well positioned to help solve. This section draws heavily on the paper by Garrow, Bodea, and Lee (2009).

As noted earlier, for all choice models based on random utility maximization theory, the total utility for alternative i (suppressing the index of n for an individual) is expressed as $U_i = \alpha_i + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_K X_{Ki} + \varepsilon_i$ where α_i is the alternative specific constant, X_{Ki} is the explanatory variable associated with the k^{th} variable, and β_K is the "true" parameter associated with explanatory variable X_{ki}.

Thus, the question of interest is how to generate the X_{ki}'s, α_i's, and β_k's for each alternative.

Generation of Systematic Utility

One method that is commonly used[3] to generate values for an explanatory variable is to use the inverse cumulative distribution function of a standard normal evaluated at one of five (or more) possible probability values. For example, given the vector $\boldsymbol{p}$ of possible probability values consisting of 0.1, 0.3, 0.5, 0.7, and 0.9 and a universal choice set consisting of six alternatives, a design matrix with 15,625 entries for a 5^6 full factorial design (i.e., a single five-level probability vector $\boldsymbol{p}_j$ for every available alternative $i, i = \overline{1,6}$) can be created and used to randomly assign values to the X_i's.

In general, as noted by Williams and Ortuzar (1982), the values for the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ parameters should be selected to ensure that none of the utility components (i.e., α_i, $\beta_i X_{ik}$, or ε_i) dominate the probability that alternative i is selected. Conceptually, this is done to ensure that a more "efficient" data generation process in the sense that the probability alternative i is selected is not entirely controlled by the value of any of the utility components, as would be the case for a total utility expression of the form $U_i = 15 + 0.5X_i + \varepsilon_i$ with X_i and ε_i distributed iid N(0,1) and iid Gumbel (0,1), respectively.

Generation of Correlation among Alternatives

Although the generation of systematic utility is focused on how to generate α_i, β_{ik}, and X_{ik} in the utility function for alternative i, the generation of correlation among alternatives is focused on how to determine the chosen alternative associated with C_n, the choice set for individual n. There are two primary methods that have been reported in the literature for accomplishing this objective. The first method that has been used to generate the chosen alternative uses P_{ni}, the probability individual n chooses alternative i, which is defined by assumptions on the error terms. For example, given four alternatives with probabilities {0.3, 0.2, 0.4, 0.1}, the chosen alternative can be assigned using a random draw. Draws with values less than 0.3 result in alternative one being chosen, values between 0.3 and 0.5 result in alternative two being chosen, values between 0.5 and 0.9 result in alternative three being chosen, and values above 0.9 result in alternative four being chosen. The second method that has been used to generate the chosen alternative simulates the error components, adds these components to V_i, and assigns the alternative with the maximum utility, defined by $U_i = V_i + \varepsilon_i$, as the chosen alternative. Note that in the context of the NL model, ε_i is defined as the total error associated with alternative i.

3 The author is grateful to Juan de Dios Ortuzar for his insights related to data generation procedures.

There are pros and cons associated with each method. The first method is simple, particularly for those models that have closed-form probability expressions such as the multinomial logit (MNL), NL, and other models belonging to the Generalized Extreme Value (GEV) family covered in Chapter 4. However, more complex models, such as the mixed logit covered in Chapter 6, have probability expressions that require numerical approximation, which may introduce a source of bias that is difficult to isolate from the potential bias introduced by the underlying data generation process. The second method is more general, but can be difficult to operationalize when error components follow a Gumbel distribution. The use of normals to approximate Gumbels is one way in which this method has been operationalized in prior studies, but as shown by Garrow, Bodea, and Lee (2009), the use of normals introduces bias in the data generation process.

Within the discrete choice modeling community, it has been common to generate the chosen alternative based on closed-form probability expressions (when available) and/or by approximating the Gumbel distribution with normals. Within the statistics community, various methods have been developed to generate Bivariate Gumbel models; however, although several studies have examined the ability of these methods to accurately recover correlation between error components (e.g., see Tiago de Oliveira 1974, 1982; Dener and Sungur 1991; Gumbel and Mustafi 1967; Shi and Zhou 1999; and Stephenson 2003), few—if any—have examined whether these Gumbel error components, when combined with the systematic portion of utility, result in unbiased parameter estimates.

Conceptually, the difficulty of generating synthetic data based on error components to replicate the desired correlation of a NL structure lies in the special construct of the NL error terms. Specifically, although the independent error components of alternatives in the same nest are considered to follow a Gumbel $G\ (0,\gamma / \mu_m)$ distribution, the common error components, which give the correlation among alternatives in the nest, are only partially specified. In essence, these components, when added to the independent ones, are assumed to generate alternative specific error components that are iid Gumbel $G\ (0,\gamma)$. Given that the distribution of the difference between two Gumbel distributions with different scale parameters does not have a parametric form (see Figure 2.8), researchers have explored several alternative approaches to generate synthetic discrete choice datasets.

One of the most common approaches is similar in spirit to methods found in the simulation literature for generating two normal random variables that are correlated (see Law and Kelton 2000: 440-48, 480-81). Conceptually, a single error term that is added to the observed utility is created. Unlike the case for a MNL model where these error terms are distributed independently and identically across alternatives, error terms are generated so that they exhibit a multivariate distribution, i.e., they are identically but not independently distributed. The multivariate distribution is generated based on a normal distribution and an approximation to a multivariate Gumbel distribution is derived by using appropriate relationships. See Garrow,

Bodea, and Lee (2009) for a detailed outline of this procedure and Garrow and Bodea (2005) for a numerical example.

Under certain conditions, it is viable to generate error components directly using Gumbels. As discussed extensively in Kotz, Balakrishnan, and Johnson (2000: 629-40), there are three primary types of Bivariate Gumbel Models (defined as Type A, Type B, and Type C). Type A, although intuitive and easily derived from univariate Gumbel distributions, has a product moment correlation that is difficult to evaluate. Type B (also known as the logistic model) is the model most frequently encountered in derivations of discrete choice models. Although the product moment correlation is easily derived from the Type B model, it is difficult to generate bivariate random variables from its cdf or pdf. Specifically, Cardell (1997) proved that the class of conjugate distributions to the Gumbel distribution exists and showed that for ε Gumbel distributed with scale parameter (γ/μ_m), $0 < \mu_m \leq 1$, $\gamma > 0$, and $\gamma / \mu_m > 0$, for v independently distributed, $(v + \varepsilon)$ is Gumbel distributed with scale parameter γ, $(\gamma < \gamma / \mu_m)$ if and only if v is $C\ (\mu_m, \mu_m / \gamma)$ distributed.[4] Although various valuable results can be obtained in closed-form using the $C\ (\gamma)$ distribution, its use is limited due to the inability to efficiently operationalize it.

In contrast, the Bivariate Gumbel Distribution (Type C or biextremal model by Tiago de Oliveira 1982) is relatively easy to operationalize. Type C has a cdf function defined as:

$$F_{X,Y}(x,y) = \exp\left[-\max\left\{e^{-x}\right\} + (1-\phi)e^{-y}, e^{-y}\right] \quad \text{for } 0 < \phi < 1$$

that is generated as the joint distribution of X and Y where:

$$Y = \max\left(X + \log(\phi), Z + \log(1-\phi)\right) \tag{3.2}$$

where X and Z are mutually independent and distributed standard Gumbel. The correlation between X and Y is given as:

$$\rho\ (\phi) = -6\pi^{-2} \int_0^{\phi} (1-t)^{-1} \log(t)\, dt \tag{3.3}$$

To generate standard bivariate Gumbel errors $(\varepsilon_{2m-1}, \varepsilon_{2m})$, $j = 1,\ldots,3$ for pairs of alternatives in the same nest having correlation $\rho_m = 1 - \mu_m^2$, the following steps are performed:

4 The probability density function of a $C(\lambda)$ distributed variable v is:

$P_\lambda(v) = \frac{1}{\lambda} \cdot \sum_{n=0}^{\infty}\left[\left((-1)^n \cdot exp(-n \cdot v)\right) / \left(n! \cdot \Gamma(-\lambda \cdot n)\right)\right]$ If v is distributed as $C(\lambda)$ and δ is a fixed scalar, then $\delta \cdot v$ is said to be distributed as $C(\lambda, \delta)$.

1. Compute the correlation $\rho_m = 1 - \mu_m^2$ of each nest. Then using Equation 3.3, find $\phi_m = \rho^{-1}(\rho_m)$.
2. For each nest, generate two independent standard Gumbel random variables $(\varepsilon_{2m-1}, z_{2m-1}) \sim G(0,1)$. By Equation 3.2, the other Gumbel component ε_{2m} in the same nest can be generated as follows:

$$\varepsilon_{2m} = \max\left(\varepsilon_{2m-1} + \log(\phi_m), z_{2m-1} + \log(1-\phi_m)\right)$$

Note that extension to multivariate Gumbels to represent general NL model structures results in product moment correlations that are more complex.

Open Research Areas Related to Generation of Synthetic Discrete Choice Datasets

Garrow, Bodea, and Lee (2009, Table 12) compare the three methods for generating chosen alternatives using design of experiments and analysis of variance methods. Specifically, for a nesting structure of six alternatives and three nests (two alternatives in each nest), they examine the ability to recover “true” parameters for 36 treatments that vary the maximum amount of correlation in a nest, the difference in correlations among nests, and the choice frequencies. The results of their analysis are summarized in Table 3.3. The method that uses nested logit probabilities to generate the chosen alternative results in unbiased parameter estimates, but is limited in the sense that it cannot be readily extended to models that have open-form probability expressions. The method that is based on Gumbel error component approximations reveals that although the error components themselves are unbiased, subtle empirical identification problems can arise when these error components are combined with synthetically generated utility functions using the procedure discussed in this chapter. The method that is based on normal error component approximations reveals that all logsum coefficients are biased upwards; the bias dramatically increases for those nests that have a low choice

Table 3.3 Pros and cons of data generation methods

Method	Pros	Cons
Probability	Simple, unbiased	Limited application
Normal	Easy to generalize	Clearly biased
Bivariate Gumbel	Unbiased variances and correlations	Care required when combining with utility; Extensions to multivariate cases needed.

Source: Adapted from Garrow, Bodea and Lee, Table 12. Reproduced with permission of Springer.

frequency and is most pronounced for those nests with high correlations among alternatives.

Several directions for future research emerge from the Garrow, Bodea, and Lee (2009) study. First, in the context of the Bivariate Gumbel distributions, future research is required to extend this method to multivariate distributions, particularly those that can be used to generate Generalized Extreme Value models that allocate alternatives to more than one nest (these models are discussed in depth in the next chapter). In this case, the structure of the Bivariate Gumbel distributions are such that the researcher will encounter the same problem observed in the context of the generalized nested logit and cross-nested logit models, namely that the maximum amount of correlation that can be accommodated in a particular nest is related to (and constrained by) the maximum amount of correlation in other nests. Second, there is a research need to determine under what conditions the Bivariate and Multivariate Gumbel distributions can be used to generate synthetic discrete choice datasets. That is, when using these distributions, researchers will need to consider both the ability of these extensions to recover variance and correlations, as well as the parameter estimates associated with the systematic portion of utility. Alternative methods for generating the systematic portion of utility may be required when using the Bivariate and Multivariate Gumbel distributions.

A Cautionary Note: Different Kinds of Nested Logit Models

The NL model presented in this chapter was motivated by utility maximization theory. Unfortunately, some of the initial software packages used to estimate NL models were based on a formulation of the NL model that is not always consistent with utility maximization. Conceptually, the difference between these two models is due to one minor difference. As described by Koppelman and Wen (1998a), in the utility-maximizing nested logit (UMNL) model, the conditional probability of choosing alternative *i* given nest *m* includes the inverse of the logsum parameter, μ_m, in the utility for each elemental alternative. In contrast, the non-normalized nested logit model (NNML) excludes this term, as shown below:

UMNL $P_{i|m} = \dfrac{e^{\left(\frac{V_i}{\mu_m}\right)}}{\sum\limits_{j \in A_m} e^{\left(\frac{V_j}{\mu_m}\right)}}$ NNML $P_{i|m} = \dfrac{e^{(V_i)}}{\sum\limits_{j \in A_m} e^{(V_j)}}$

This apparently small difference leads to very different models. In their paper, Koppelman and Wen (1998a) present an empirical analysis that shows how different parameter estimates, nesting structures, and overall model interpretations differ when using the UMNL and NNML model.

In practice, the analyst must be very careful to know which type of NL model is being estimated by a software package. Even today, it is often difficult to discern whether the "default" estimation in a software package is based on the UMNL or NNML formulation. Some software estimate only the NNML model, others estimate only the UNML formulation, and some do both. Given a software package that only estimates NNML models, it is possible to "trick" the software into estimating a UMNL model by effectively adding levels to a tree and appropriately constraining logsum coefficients across nests—this trick is explained in detail in the Koppelman and Wen (1998a) paper.

To summarize, it is the author's opinion that the UMNL model should always be used over the NNML model due to its stronger foundation in behavioral theory and clear and intuitive relationships to substitution patterns among alternatives. Unfortunately, based on the author's experiences, it appears that SAS (the statistical software that is commonly used in the airline industry) estimates only the NNML model and does not currently provide an add-on module that can be used to estimate the UMNL model. Thus, an analyst using SAS who wants to estimate a UMNL model would need to use the "trick" referenced above to estimate NL models. It is important to note, though, that the use of the trick (which involves constrained optimization) is not as efficient as using unconstrained optimization to directly solve for the parameters of the UMNL model. In addition, using the "trick" becomes practically infeasible as the number of levels in the nesting structure grows. Consequently, it is the author's opinion that airlines embarking on the estimation of NL models should carefully evaluate different software packages and perform some initial tests to: 1) ensure the estimates returned are based on utility-maximization theory; and, 2) verify that the software can handle a large number of observations and/or nesting structures with multiple levels.

Not to complicate matters more, but in the airline industry there is yet another point of confusion related to "nested logit models" that appears in discussions of itinerary choice models. This point is extensively discussed in Chapter 7. The bottom line, however, is that when working with nested logit models, analysts should be aware that there are many subtle differences, often not clearly documented, that they may encounter. Analysts who understand how NL probabilities presented in this chapter are calculated can use these formulas to verify whether the software they are using is based on a utility-maximizing framework.

Summary of Main Concepts

This chapter presented fundamental concepts related to the theory and estimation of NL models. The most important concepts covered in this chapter include the following:

- The NL model relaxes the independence of irrelevant alternatives (IIA) property of the MNL model by allowing error components to be correlated. In a NL model, alternatives belong to one and only one nest. Alternatives that belong to the same nest share a common error term (or covariance).

Alternatives that belong to different nests have independent error terms; this property is referred to as the independence of irrelevant nests (IIN). The IIN property can be seen by the fact that NL cross-elasticities for alternatives in different nests are identical to MNL cross-elasticities.

- Correlation is a measure of the amount of substitution or competition among alternatives in the nest. High correlation leads to greater competition effects among alternatives in the nest, i.e., an improvement in alternative i in nest m will draw proportionately more share from alternatives in the nest. Formally, correlation in a nest is the ratio of common variance to total variance, or $\rho^2 = \left(1 - \mu_m^2\right)$
- In order to be consistent with utility maximization theory, logsum coefficients associated with a nest range from (0,1]. Values closer to zero indicate more correlation among alternatives in the nest whereas values closer to one indicate less correlation. A value of $\mu_m = 1$ for all nests is equivalent to a MNL.
- In three-level NL models, logsum coefficients must decrease as one moves down the tree. This is to ensure the model is consistent with utility maximization, i.e., an improvement in one alternative does not lead to a decrease in the probability that alternative is chosen. Alternatives in lower level nests exhibit higher correlations (and more competition) than alternatives in higher level nests.
- The theoretical derivation of the NL model is based on two assumptions: 1) the total variance of an alternative is identically distributed $G(0,\gamma)$; and, 2) the variance of the independent component of an alternative in nest m is distributed $G(0,\gamma / \mu_m)$. The distribution of the common variance is given as the difference between two Gumbels with different scales, which poses challenges in creating unbiased simulated NL datasets.
- NL choice probabilities can be derived as the product of conditional and marginal probabilities. The conditional probability is given as the probability of selecting alternative i among all j alternatives in nest m, conditional on the choice of m, and the marginal probability is the probability of selecting nest m among all nests. This formulation is particularly helpful when extending NL models to include additional levels of nests.
- Analysts must use care when using off-the-shelf software to estimate NL models, as many are based on formulations that are not consistent with utility maximization.

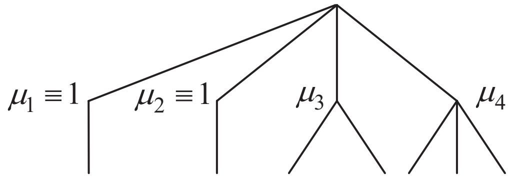


Figure 3.4 Notation for a two-level NL model

Appendix 3.1: Derivation of the NL Model

This section derives choice probabilities for a two-level nested logit model; the extension to multi-level NL models is straightforward. Figure 3.4 above shows a general representation of a two-level NL model that contains four nests and seven alternatives. The total number of alternatives in the choice set, J, is partitioned into M non-overlapping nests $A_1, A_2,\ldots,A_m$ where $i \in A_m$ is the set of alternatives belonging to nest m. Note alternative i can be assigned to only one nest. A "degenerate nest" is defined as a nest that contains only one alternative. The logsum coefficient associated with a degenerate nest is defined to be one.

The derivation of the NL probabilities uses the fact that the probability of choosing alternative i can be expressed as the product of conditional and marginal probabilities: $P_i = (P_i \mid m) \times P_m$. The first component of the product is the probability of selecting alternative i among all J alternatives in nest m, conditional on the choice of m, and the second component of the product is the marginal probability of selecting nest m among all nests.

Given a conceptual overview of the derivation, formal notation and definitions are provided, followed by the steps of the derivation. To simplify the notation, the index for individual n has been suppressed. Formally, utility is defined as:

$$U_i = V_i + V_m + \varepsilon_i + \varepsilon_m$$

where:

U_i Total utility for the i^{th} alternative,
V_m Deterministic, or observed, utility for the i^{th} alternative that is common to all alternatives in nest m,
V_i Deterministic, or observed, utility associated with the i^{th} alternative,
ε_m Error associated with the m^{th} nest (common to all alternatives in nest m),
ε_i Error associated with the i^{th} alternative.

The total utility associated with nest m, U_m, is given by the maximum of the utilities in that nest. That is,

$$U_m = \max_{j \in A_m} \left(U_j \right)$$

The NL model is defined via assumptions on the error terms. Formally:

$$\varepsilon_i \sim G\left(0, \frac{1}{\mu_m}\right), \quad 0 < \mu_m \leq 1, \quad \Rightarrow \text{var} = \frac{\pi^2}{6\left(\frac{1}{\mu_m}\right)^2}$$

$$\varepsilon_i + \varepsilon_m \sim G(0,1) \Rightarrow \text{var} = \frac{\pi^2}{6}$$

Note that μ_m is bounded between zero and one so that the variance associated with the distinct error components is less than or equal to the total error variance. Although it has been shown that values above one are theoretically possible (e.g., see Kling and Herriges 1995; Herriges and Kling 1996; Train McFadden and Ben-Akiva 1987a), for practical purposes (and without loss of generality for this proof), it is assumed $\mu_m \in (0,1)$.

The probability of selecting alternative i among all J alternatives in nest m, conditional on the choice of m is given as:

$$P_i \mid m = P\left(U_i > U_j\right) \quad \forall j \in A_m, \; i \neq j$$

$$= P\left(V_i + V_m + \varepsilon_i + \varepsilon_m > V_j + V_m + \varepsilon_j + \varepsilon_m\right) \quad \forall j \in A_m, \; i \neq j$$

$$= P\left(\varepsilon_j \leq V_i - V_j + \varepsilon_i\right) \quad \forall j \in A_m, \; i \neq j$$

Note that this expression is similar to that obtained when deriving MNL probabilities (shown in Appendix 2.1). However, the MNL model assumes $\varepsilon_j \sim G\,(0,1)$, $\forall\, j \in C_n$ whereas the NL model assumes $\varepsilon_j \sim G\,(0,1/\mu_m)$, $\forall\, j \in A_m$. Therefore, the expression obtained for the MNL probability can be used for $P_i \mid m$ once the NL utility has been rescaled by dividing by μ_m so that $\varepsilon_j^* \sim G(0,1)$, $\forall j \in A_m$. Keep in mind that since μ_m is defined as the "inverse scale," dividing by μ_m implies the variance of ε_j^* is $\pi^2/6$.

$$P\left(\varepsilon_j \leq V_i - V_j + \varepsilon_i\right) \qquad ; \; \varepsilon_j \sim G\left(0, \frac{1}{\mu_m}\right)$$

$$= P\left(\frac{\varepsilon_j}{\mu_m} \leq \frac{V_i}{\mu_m} - \frac{V_j}{\mu_m} + \frac{\varepsilon_i}{\mu_m}\right) \qquad ; \; \frac{\varepsilon_j}{\mu_m} = \varepsilon_j^* \sim G(0,1)$$

Using the result from the MNL derivation, the conditional probability becomes:

$$P_i \mid m = \frac{e^{\left(\frac{V_i}{\mu_m}\right)}}{\sum_{j \in A_m} e^{\left(\frac{V_j}{\mu_m}\right)}}$$

The marginal probability of selecting nest m among all M nests is given as:

$$P_m = P\left(U_m > U_l\right) \quad \text{for } \; l = 1, 2, \ldots, M; \; l \neq m$$

The derivation of P_m uses the fact that the total utility associated with a nest is given by the maximum of the utilities associated with the alternatives in that nest. Specifically, P_m is derived by using a property of the Gumbel distribution discussed in Chapter 2. This property states that given $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_J$ are independently distributed Gumbel such that they have the same scale, but different modes, the distribution of their maximum is:

$$\max_j\left(\varepsilon_j\right) \sim G\left(\frac{1}{\gamma} \ln \sum_{j=1}^{J} \exp\left(\gamma \eta_j\right), \; \gamma\right)$$

Applying this property to the total utility associated with nest l gives, for $l = 1,2,\ldots,M$, $l \neq m$:

$$U_l = \max_{j \in A_l}\left(U_j\right) = \max_{j \in A_l}\left(V_j + V_l + \varepsilon_j + \varepsilon_l\right)$$

$$U_l = V_l + \varepsilon_l + \max_{j \in A_l}\left(V_j + \varepsilon_j\right); \quad \varepsilon_j \sim G\left(0, \frac{1}{\mu_l}\right)$$

$$U_l = V_l + \varepsilon_l + G\left(\mu_l \Gamma_l, \frac{1}{\mu_l}\right) \; \text{where } \; \Gamma_l = \ln\left(\sum_{j \in A_l} e^{\left(\frac{V_j}{\mu_l}\right)}\right)$$

The probability of selecting nest m can be rewritten for $l = 1,2,\ldots,M$, $l \neq m$:

$$P_m = P\left(V_m + \varepsilon_m + \mu_m \Gamma_m + \varepsilon_{i'}^{*} > V_l + \varepsilon_l + \mu_l \Gamma_l + \varepsilon_{j'}^{*}\right), \; \varepsilon_{i'} \sim G\left(0, \frac{1}{\mu_m}\right); \varepsilon_{j'} \sim G\left(0, \frac{1}{\mu_l}\right)$$

$$P_m = P\left(\varepsilon^*_{j'} + \varepsilon_l \le (V_m + \mu_m \Gamma_m) - (V_l + \mu_l \Gamma_l) + \left(\varepsilon_m + \varepsilon^*_{i'}\right)\right)$$

By definition, $\left(\varepsilon^*_{j'} + \varepsilon_l\right), \left(\varepsilon^*_{i'} + \varepsilon_m\right) \sim G(0,1)$ and the result from the MNL derivation can be applied:

$$P_m = \frac{e^{V_m + \mu_m \Gamma_m}}{\sum_{l=1}^{M} e^{V_l + \mu_l \Gamma_l}}$$

In summary, the probability of choosing alternative *i* is given as:

$$P_i = (P_i \mid m) \times P_m$$

$$P_i = \frac{e^{\left(\frac{V_i}{\mu_m}\right)}}{\sum_{j \in A_m} e^{\left(\frac{V_j}{\mu_m}\right)}} \times \frac{e^{V_m + \mu_m \Gamma_m}}{\sum_{l=1}^{M} e^{V_l + \mu_l \Gamma_l}}, \Gamma_m = \ln\left(\sum_{j \in A_m} e^{\left(\frac{V_j}{\mu_m}\right)}\right)$$

Appendix 3.2: Derivation of NL Correlations

This section derives the correlation between alternatives *i* and *j* that share the same nest in a NL model. The logsum parameter associated with this nest is given as μ_m. Utility for alternatives *i* and *j* are given as:

$U_k = V_k + V_m + \varepsilon_k + \varepsilon_m$ for $k = i, j$

where:

U_k Total utility for the k^{th} alternative,
V_m Deterministic, or observed, utility for the k^{th} alternative that is common to all alternatives in nest *m*,
V_k Deterministic, or observed, utility associated with the k^{th} alternative,
ε_m Error associated with the m^{th} nest (common to all alternatives in nest *m*),
ε_k Error associated with the k^{th} alternative.

The following assumptions are imposed on error components:

$$\varepsilon_k \overset{IID}{\sim} G\left(0, \frac{1}{\mu_m}\right) \quad 0 < \mu_m \le 1, \quad \Rightarrow \text{var} = \frac{\pi^2}{6\left(\frac{1}{\mu_m}\right)^2}$$

By definition, ε_i and ε_j are independent. Conceptually, for $k = i, j$, ε_k and ε_m are also independent because the "error" common to both alternatives, reflected in ε_m, is not influenced by the "error" associated just with alternative *k*. The correlation between alternatives *i* and *j* is given as a function of the covariance and variance terms associated with these alternatives. Formally:

$$COV\left[U_i, U_j\right] = COV\left[\varepsilon_i + \varepsilon_m, \varepsilon_j + \varepsilon_m\right]$$

$$COV\left[U_i, U_j\right] = COV\left[\varepsilon_i, \varepsilon_j\right] + COV\left[\varepsilon_i, \varepsilon_m\right] + COV\left[\varepsilon_m, \varepsilon_j\right] + COV\left[\varepsilon_m, \varepsilon_m\right]$$

$$COV\left[U_i, U_j\right] = COV\left[\varepsilon_m, \varepsilon_m\right] = V\left[\varepsilon_m\right]$$

$$V\left[\varepsilon_i + \varepsilon_m\right] = V\left[\varepsilon_i\right] + V\left[\varepsilon_m\right] \quad \rightarrow \quad V\left[\varepsilon_m\right] = V\left[\varepsilon_i + \varepsilon_m\right] - V\left[\varepsilon_i\right]$$

$$V\left[\varepsilon_m\right] = \frac{\pi^2}{6} - \frac{\pi^2}{6} \cdot \mu_m^2 = \frac{\pi^2}{6} \cdot \left(1 - \mu_m^2\right)$$

$$corr\left[U_i, U_j\right] = \frac{COV\left[U_i, U_j\right]}{\sqrt{V\left[U_i\right] \cdot V\left[U_j\right]}} = \frac{V\left[\varepsilon_m\right]}{\sqrt{V\left[\varepsilon_i + \varepsilon_m\right] \cdot V\left[\varepsilon_j + \varepsilon_m\right]}}$$

$$corr\left[U_i, U_j\right] = \frac{\frac{\pi^2}{6} \cdot \left(1 - \mu_m^2\right)}{\frac{\pi^2}{6}} = \left(1 - \mu_m^2\right)$$

Chapter 4
Structured Extensions of MNL and NL Discrete Choice Models

Laurie A. Garrow, Frank S. Koppelman, and Misuk Lee

Introduction

Although the MNL and NL models are most frequently used in practice, dozens of other discrete choice models are available, several that are particularly relevant for airline applications. From a historical perspective, the development of discrete choice models evolved along two parallel streams of research. This is due in part to the formulation of the multinomial probit model (MNP) by Daganzo in 1979, which appeared at approximately the same time as the MNL and NL models. Theoretically, the derivation of the MNL, NL, and MNP models arise due to different assumptions imposed on error terms. Although the assumption that the error terms are iid G(0,1) leads to the elegant, yet restrictive MNL model; the assumption that the error terms follow a multivariate normal distribution leads to the MNP. Unlike the MNL, the probit model allows flexible substitution patterns, correlation among unobserved factors, heteroscedasticity, and random taste variation. However, the choice probabilities can no longer be expressed analytically in closed-form and must be numerically evaluated. In practical terms, it has been difficult to use the probit in applications that require the numerical evaluation of more than approximately ten integrals.

Conceptually, MNL and MNP models can be loosely thought of as the endpoints on a spectrum of discrete choice models. On one end is the MNL, a restrictive model that has a closed-form probability expression that is computationally simple. On the other end is the MNP, a flexible model that has a probability expression that must be numerically evaluated. Since the 1970's, advances in discrete choice models have generally focused on either relaxing the substitution restriction of the MNL while maintaining a closed-form expression for the choice probabilities (such as the NL model) or reducing the computational requirements of open-form models and further expanding the spectrum of open-form models to include more general formulations.

This chapter has two primary objectives. The first is to provide an overview of the development of different discrete choice models that occurred after the appearance of the MNL, NL, and MNP models. A specific emphasis is placed on highlighting those models that are most relevant from either a theoretical context or from an aviation applications context. The second objective is to provide an

in-depth examination of models that fall within the class of Generalized Extreme Value (GEV) models, namely two-level models that belong to the generalized nested logit (GNL) family and multi-level models that belong to the Network GEV family. The MNL and NL models, covered in earlier chapters, fall within the GEV class. However, the GEV class contains numerous other models that relax the independence of irrelevant alternatives (IIA) property associated with the MNL or the independence of irrelevant nests (IIN) property associated with the NL by allowing alternatives to be allocated to more than one nest. From a theoretical perspective, the GEV class of models is very powerful, as it provides researchers with a general framework they can use to create a discrete choice model that will be consistent with random utility theory. From a practical perspective, GEV models are particularly relevant to itinerary choice problems, where substitution among alternatives commonly occurs simultaneously along multiple dimensions (e.g., carrier, level of service, time of day).

To help visualize differences across GEV models, an appendix is included at the end of this chapter that summarizes probabilities, direct-elasticities, and cross-elasticities for the GEV models discussed in this chapter. This chapter draws heavily on the work by Sethi (2000), Koppelman and Sethi (2000), Coldren (2005), Coldren and Koppelman (2005a, 2005b), and Koppelman (2008).

Historical Development of Discrete Choice Models

The proliferation of discrete choice models developed since the 1970's is evident in Figure 4.1, which classifies these developments according to how they relax the assumptions of the MNL model. The motivation for tracing the development of dozens of discrete choice models shown in Figure 4.1 is not to provide a comprehensive treatment of each model, but rather to underscore the fact that the development of discrete choice models has been a very fruitful area of research, and goes beyond the simple MNL, NL, and probit models that are often the only models that are covered in traditional operations research departments (where most airline practitioners conduct their undergraduate and/or graduate work). Although the airline industry is becoming more comfortable with using discrete choice models (and specifically MNL and NL models) for customer choice applications, there are many opportunities to better leverage these models. The goal of introducing the wide range of models available for airline applications is to help expand the focus of current research, mainly limited to integrating simplistic MNL formulations within advanced optimization algorithms. For the potential of discrete choice models to be fully realized within the airline industry, more sophisticated specifications and more advanced discrete choice models than those currently used will be required (particularly within the revenue management area). There is a genuine need to balance the integration of sophisticated optimization techniques with realistic choice models that capture the underlying behavior of customers; failure to achieve sophistication on both the optimization and discrete

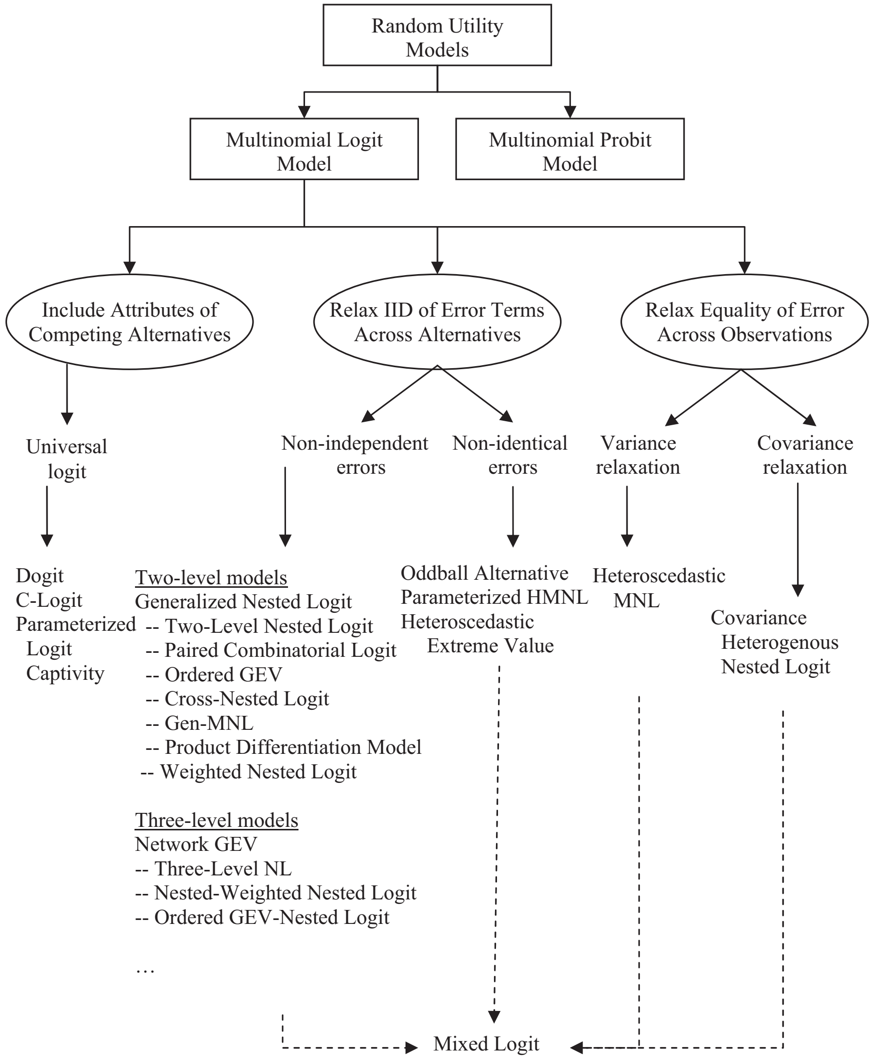


Figure 4.1 Overview of the origin of different logit models

Source: Adapted from Sethi 2000: Exhibit 2.1 (reproduced with permission of author).

choice sides of the equation is likely to lead to limited success and limited revenue gains.

With this motivation as background, Figure 4.1 traces the evolution of discrete choice models. The models portrayed in the figure are not an exhaustive list of all discrete choice models, but are representative. Figure 4.2 presents an alternative classification of these developments according to the time they appeared in the

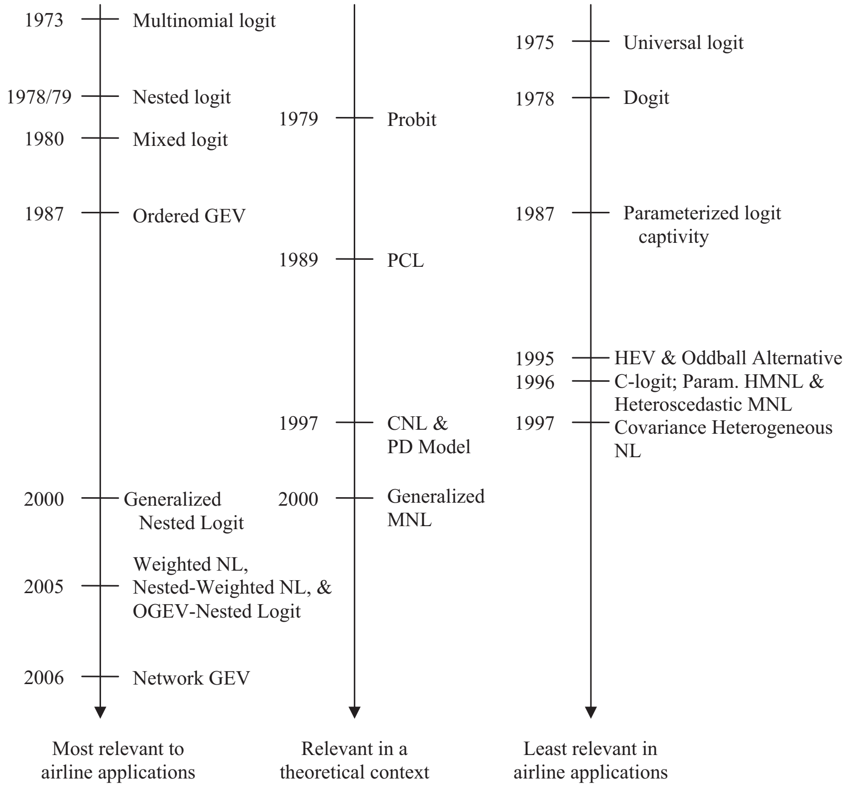


Figure 4.2 Classification of logit models according to relevance to the airline industry

literature and an assessment of their relevance to the airline industry (either from practical or theoretical perspectives).

Relaxations of the MNL have occurred along three lines of development. The universal logit model (also called the mother logit model) was proposed early on to relax the IIA assumption of the MNL model by including attributes of competing alternatives in the utility function for each alternative (McFadden 1975). Extensions of the universal logit model tailored to specific applications include the Dogit model (Gaudry and Dagenais 1978), the Parameterized Logit Captivity model (Swait and Ben-Akiva 1987), and the C-Logit model (Cascetta Nuzzola Russo and Vitetta 1996). The Dogit model incorporates a "captivity parameter" to reflect the resistance of individuals to switch from certain alternatives, e.g., itinerary choice models: individuals based in Atlanta may be loyal to Delta Air Lines and unwilling to consider Air Tran alternatives when selecting their itineraries. The Parameterized Logit Captivity model is a relaxation of the Dogit model in the sense that it allows this captivity parameter to be estimated as a non-negative function of decision-maker and alternative characteristics; conceptually, this results in a probabilistic

choice set formulation. The C-Logit model, used in route choice models, accounts for non-independence in routes by incorporating a similarity index that represents the amount of route overlap. To the authors' knowledge, no applications of the Universal Logit, Dogit, Parameterized Logit Captivity, or C-Logit models exist in the aviation industry. As summarized by Koppelman and Sethi (2000), this "may be due to lack of consistency with utility maximization in some cases, the potential to obtain counter-intuitive elasticities, and the complexity of search for a preferred specification (Ben-Akiva 1974)."

A second line of development with relaxing the MNL assumptions was to relax the assumption that the total variance associated with an alternative is identically distributed and/or that the covariance across alternatives is identically distributed. The Heteroscedastic MNL model, proposed by Swait and Adamowicz (1996) is an example of a model that relaxes the assumption of identical variance across alternatives. Within transportation, there are numerous situations that arise where some alternatives are expected to exhibit higher variance than other alternatives. Many of these applications arise in the context of stated preference surveys, e.g., it may be desirable to model respondent fatigue (represented in less precise and/or more rushed answers later in the survey). However, there are also contexts based on revealed preference data that arise in broad transportation contexts, such as in highway route choice models where the variance associated with travel times may increase with distance. Within the aviation context, it is less clear whether this differential error variance relationship would be applicable, as trip distance is closely tied with equipment type, domestic versus international trips, mix of mainline versus connecting/regional carriers, etc. Within transportation, there are also numerous situations that arise where some alternatives are expected to exhibit different correlations. The Covariance Heterogeneous Nested Logit model by Bhat (1997) relaxes the assumption of equal correlation across alternatives by parameterizing the logsum coefficient as a function of individual and trip-related characteristics. This model may be useful in some aviation contexts, albeit the limited availability of individual and trip-related characteristics could restrict the number of applications for which this model will be useful.

The final line of development was to relax the assumptions that error terms are independent and identically distributed across alternatives. As seen in Figure 4.1, this is where the majority of research efforts have been focused. The Oddball Alternative model (Recker 1995), Parameterized Heteroscedastic MNL model (Swait and Stacey 1996), and Heteroscedastic Extreme Value (HEV) model (Bhat 1995) are three examples of models that relax the assumption that errors are identically distributed. The first two models are able to incorporate heteroscedastic error terms while maintaining a closed-form probability expressions; however, the ability to have a closed-form probability is derived by imposing restrictive assumptions on the relationships among error components, which are likely to be inappropriate in many situations. In contrast, the HEV model allows error terms to be non-identically distributed across alternatives, albeit this ability results in the need to numerically evaluate probabilities (as is the case for probit models).

Among these three models, the HEV is the one that has been most commonly applied in transportation contexts (e.g., see Allenby and Ginter 1995; Hensher 1998; Hensher Louviere and Swait 1999). One of the primary benefits of this model, noted by Hensher (1998), is that it can help uncover an appropriate nesting structure (thereby eliminating the need to conduct an exhaustive search for the nesting structure that has the best fit).

In contrast to these three models that allow errors to be non-identical, dozens of models have been developed that relax the assumption that error terms are non-independent. Conceptually, these models (in addition to others) relax the independence assumption by including covariance terms that are created through allocating alternatives to two or more nests while maintaining the assumption that total variance is identically distributed across alternatives. From a theoretical perspective, these two requirements (more general covariance structure combined with equality of total variance) relax the IIA property while maintaining closed-form expressions for probabilities. However, these two requirements also impose bounds (either explicitly or implicitly) on the amount of correlation that can be incorporated between alternatives. As a consequence, the calculation of correlation among alternatives becomes much more complex and may result in the loss of closed-form formulas. In addition, it becomes critical to ensure that the number of covariance terms included in the model results in an identified model, that is, it is important to ensure that the differences between utilities of any pair of alternatives are uniquely identified.

For these reasons, the development of GEV models is accompanied by an assessment of how the IIA property is relaxed (via deriving elasticity and cross-elasticity functions), theoretical and/or empirical assessment of bounds associated with the covariance structure (via examining the maximum amount of correlation that can be incorporated), and development of identification rules. Identification rules are used to ensure the resulting model is properly normalized (that is, to ensure that differences in utility are uniquely determined). It is also common to explore empirically the properties of these models using both simulated data and datasets from practice. Conceptually, this is because empirical identification problems can arise when some alternatives are infrequently chosen or when the number of alternatives, nests, allocation parameters, and/or covariance terms becomes large. In practice, it is common to impose constraints on the relationships among logsum parameters and/or on the relationships among allocation parameters to avoid empirical identification issues that arise when using datasets from practice.

Given this overview of the theoretical and practical challenges that arise when using these more flexible models, the next two sections provide a detailed discussion of GEV models that allocate alternatives to more than one nest. It is useful to further classify these models according to whether they contain two levels or more than two levels. Two-level models belong to the family of generalized nested logit (GNL) models whereas models that contain more than three levels belong to the family of Network GEV (NetGEV) models. Although the GNL is

a special case of the NetGEV model, the distinction will be maintained, as many of the current theoretical and empirical results have been investigated in the GNL context, but not the more general (and much more recent) NetGEV context. Two-level models are reviewed first. From an aviation applications perspective, these models served as a pre-cursor to "weighted" GEV models that have been used for itinerary choice problems; some of these weighted models are presented in the second section.

Generalized Nested Logit: Two-level Models that Allocate Alternatives to More than One Nest

Several GEV models have been reported in the literature that allocate alternatives to more than one nest. Table 4.1 classifies five of these models that appeared in the literature from 1987 to 2000. As shown in the table, these models differ in how their nesting structures are defined, how they allocate alternatives among nests, and which parameters are constrained to be equal. This section discusses each of these models. A particular emphasis is placed on illustrating how choice probabilities, variance-covariance matrices, direct-elasticities, and cross-elasticities differ across these models.

Table 4.1 Comparison of two-level GEV models that allocate alternatives to nests

Model	Nesting Structure	Allocation of Alternatives to Nests	Logsum Coefficients
PCL	Defined	Constrained to be equal	Estimated
OGEV	Defined	Estimated	Estimated
CNL	General	Estimated	Constrained to be equal
Gen-MNL	General	Constrained to be equal	Estimated
GNL	General	Estimated	Estimated

Paired Combinatorial Logit

The paired combinatorial logit (PCL) model (Chu 1989; Koppelman and Wen 1998b) is one of the simplest GEV models that allocates alternatives to different nests. As shown in Figure 4.3, the PCL model allocates alternatives to multiple nests so that each pair of alternatives appears in one nest. Formally, given N alternatives, each alternative will appear in $(N-1)$ nests, implying an allocation parameter of $\tau = 1/(N-1)$. Thus, since there are four alternatives in Figure 4.3, each alternative

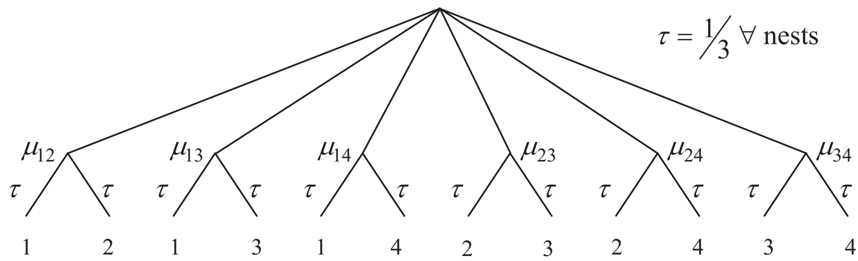


Figure 4.3 Paired combinatorial logit model with four alternatives

appears in three nests, resulting in an allocation parameter of 1/3. The total number of nests is given by N!/ {N – 1)! × 2!} = 4!/ {3!2!} = 6. Unlike allocation parameters that are defined to be equal, separate logsum coefficients are associated with each nest.

Suppressing the index for individual *n* for notational convenience, the probability of choosing alternative *i* in a PCL model is given as:

$$P_i = \sum_{j \neq i} P_{i|ij} \times P_{ij}$$

$$= \sum_{j \neq i} \left[\frac{\left(\tau e^{V_i}\right)^{\frac{1}{\mu_{ij}}}}{\left(\tau e^{V_i}\right)^{\frac{1}{\mu_{ij}}} + \left(\tau e^{V_j}\right)^{\frac{1}{\mu_{ij}}}} \times \frac{\left(\left(\tau e^{V_i}\right)^{\frac{1}{\mu_{ij}}} + \left(\tau e^{V_j}\right)^{\frac{1}{\mu_{ij}}}\right)^{\mu_{ij}}}{\sum_{r=1}^{J-1} \sum_{s=r+1}^{J} \left(\left(\tau e^{V_r}\right)^{\frac{1}{\mu_{rs}}} + \left(\tau e^{V_s}\right)^{\frac{1}{\mu_{rs}}}\right)^{\mu_{rs}}} \right], \quad 0 < \mu_{ij}, \mu_{rs} \leq 1$$

where:

μ_{ij} is the logsum coefficient associated with the nest that contains alternatives *i* and *j*,

τ is an allocation parameter that characterizes the portion of alternative *i* assigned to a nest. For the PCL model, $\tau = 1/(N-1)$ where *N* is the number of alternatives,

r,s are indices used to sum over all possible nests.

The first component of the product is the probability of selecting alternative *i* among the pair of alternatives *i* and *j*, conditional on choosing the nest that contains the pair. The second product is the probability of selecting nest *ij* among all nests. The total probability for alternative *i* is now obtained by summing over all nests that contain alternative *i*. Similar to conditions observed with the MNL and NL model, the logsum coefficients must range from 0 to 1 to ensure that the model is consistent with utility maximization. The allocation parameter, τ, can be dropped

from the PCL equation; however, it is retained to highlight key areas of distinction among the PCL and other models.

The variance-covariance matrix associated with the four-alternative PCL model is:

$$\Omega = \begin{array}{cc} & \begin{array}{cccc} \quad 1 \quad & \quad 2 \quad & \quad 3 \quad & \quad 4 \quad \end{array} \\ \begin{array}{c} 1 \\ \\ 2 \\ \\ 3 \\ \\ 4 \end{array} & \begin{bmatrix} 1 & \tau\left(1-\mu_{12}^2\right) & \tau\left(1-\mu_{13}^2\right) & \tau\left(1-\mu_{14}^2\right) \\ & & & \\ \tau\left(1-\mu_{12}^2\right) & 1 & \tau\left(1-\mu_{23}^2\right) & \tau\left(1-\mu_{24}^2\right) \\ & & & \\ \tau\left(1-\mu_{13}^2\right) & \tau\left(1-\mu_{23}^2\right) & 1 & \tau\left(1-\mu_{34}^2\right) \\ & & & \\ \tau\left(1-\mu_{14}^2\right) & \tau\left(1-\mu_{24}^2\right) & \tau\left(1-\mu_{34}^2\right) & 1 \end{bmatrix} \end{array} \times \frac{\pi^2}{6\gamma^2}$$

For notational convenience, the common term of $\pi^2/6\gamma^2$ has been factored out. Also, to facilitate comparison with more complex models, a "full" variance-covariance matrix has been defined. The upper triangular of the matrix defines the covariance between alternatives i and j in nest ij that is "weighted" by how much of alternative i is allocated to nest ij (the allocation of alternative i is represented for rows $(i, j)\ \forall\ i < j$)). Similarly, the lower triangular of the matrix defines the covariances between alternatives i and j in nest ij that is "weighted" by how much of alternative j is allocated to nest ij (the allocation of alternative j is represented for rows $(i, j)\ \forall\ i > j$)). In the case of the PCL model, the upper triangular and lower triangular are symmetric and calculations of correlations and covariance terms are straightforward. That is, the PCL allocation parameter effectively limits the maximum implied correlation in any nest to $1/(J - 1)$. (This result is obtained when $\mu_{ij} \rightarrow 0$ for all ij nests.) Thus, as the number of alternatives grows, the ability to incorporate a high degree of correlation between a pair of alternatives decreases. The direct- and cross-elasticity equations for the PCL model are given as:

$$\text{PCL direct-elasticity} = \left[\left(1-P_i\right)+\sum_{j\neq i}\left(\frac{1-\mu_{ij}}{\mu_{ij}}\right)\frac{P_{i|ij}P_{j|ij}P_{ij}}{P_i}\right]\beta_k X_{ik}$$

$$\text{PCL cross-elasticity} = -\left[P_i+\left(\frac{1-\mu_{ij}}{\mu_{ij}}\right)\frac{P_{i|ij}P_{j|ij}P_{ij}}{P_j}\right]\beta_k X_{ik}$$

From an aviation applications perspective, there are few (if any) contexts in which the PCL model would be used. However, from a theoretical perspective, the PCL model is highly relevant, as it was one of the first models of the GEV class to incorporate a general variance-covariance matrix in which all covariance terms were positive. As noted earlier, covariance terms must be positive to ensure that when an improvement is made to alternative i, it will draw proportationately

more share from alternatives that share the same nest; a negative covariance would imply less competition among alternatives that share the same nest. The PCL model, although simple, is also helpful for highlighting why there are limits on the maximum amount of correlation that can be incorporated among alternatives.

Ordered Generalized Extreme Value

The Ordered Generalized Extreme Value (OGEV) model (Small 1987) is similar to the PCL model in the sense that the nesting structure is defined by the model. However, instead of defining nests for every possible pair of alternatives, the OGEV model is used in applications in which the ordering of alternatives has a physical meaning. For example, the OGEV model can be used to capture time of day competition effects among airline itineraries. Figure 4.4 shows an OGEV model for six alternatives ($J = 6$) and one adjacent time period ($T = 1$). In contrast to notation used thus far, note that nest one is not defined as the first nest on the left, but rather the first non-degenerate nest, i.e., the first nest that contains more than one alternative. Also, note that the first and last nests are degenerate in that there is only one alternative in each of these nests; since the logsum parameter is not identified for degenerate nests, it is commonly set to one. The total number of nests is given as (J-T+$2T$) where (J-T) is the number of nests that contain more than one alternative and $2T$ is the number of nests that contain one alternative.

The OGEV probability is given as:

$$P_i = \sum_{m=i}^{i+T} P_{i|m} \times P_m$$

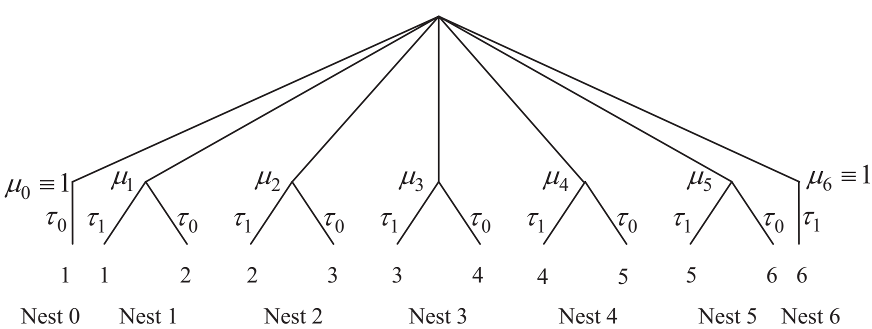


Figure 4.4 Ordered GEV model with one adjacent time period

$$= \sum_{m=i}^{i+T}\left[\frac{\left(\tau_{m-i}e^{V_i}\right)^{\frac{1}{\mu_m}}}{\sum_{j\in A_m}\left(\tau_{m-j}e^{V_j}\right)^{\frac{1}{\mu_m}}}\times\frac{\left(\sum_{j\in A_m}\tau_{m-j}e^{\frac{V_j}{\mu_m}}\right)^{\mu_m}}{\sum_{r=1}^{J+T}\left(\sum_{s\in A_r}\tau_{r-s}e^{\frac{V_s}{\mu_r}}\right)^{\mu_r}}\right],\quad 0<\mu_m\le 1;\quad \sum_{m=1}^{J+T}\tau_m=1$$

where:

T	is the number of adjacent time periods in the OGEV model,
J	is the total number of alternatives,
$j \in A_m$	is the set of all alternatives that belong to nest m,
r	is an index used to sum over all nests,
τ_{m-i}	are unknown allocation parameters that characterize the portion of alternative i assigned to a nest. Allocation parameters are non-negative, i.e., $\tau_{m-i} \ge 0$ and must sum to one for every alternative. Defining nests for a T-step OGEV model as shown in Figures 4.4 and 4.5, alternative i belongs to nests i-1, i, i+1, …, $i+T$, this last condition is equivalent to $\sum_{m=1}^{J+T}\tau_m = 1$
μ_m	is the logsum coefficient associated with nest m, m=1,…, $J+T$.

The first component of the product is the probability of selecting alternative i among all alternatives that belong to nest m, conditional on choosing nest m. The second product is the probability of selecting nest m among all nests. Consistent with the PCL model, the total probability for alternative i is obtained by summing over all nests that contain alternative i. Distinct from the PCL model, the portion of an alternative that shares a nest with itineraries that depart in the time period immediately before (τ_0) or the time period immediate after (τ_1) is estimated from the data. A constraint is also added to ensure that $\tau_0 + \tau_1 = 1$.

From an interpretation perspective, a value of $0.5 < \tau_0 < 1$ (and $0 < \tau_1 < 0.5$) means that an itinerary departing in time period three would compete more with itineraries in the earlier time period two than with itineraries in the later time period four. Intuitively, this result would be expected for outbound itineraries for travelers that need to arrive at their destinations by a fixed time. The increased substitution among alternatives that depart in the same time period or adjacent time periods is also seen in the direct-elasticity and cross-elasticity equations:

$$\text{OGEV direct-elasticity} = \left[\left(1-P_i\right)+\sum_{m=i}^{i+T}\left(\frac{1-\mu_m}{\mu_m}\right)\frac{P_{i|m}P_m\left(1-P_{i|m}\right)}{P_i}\right]\beta_k X_{ik}$$

$$\text{OGEV cross-elasticity} = \left[P_i+\sum_{m=i}^{i+T}\left(\frac{1-\mu_m}{\mu_m}\right)\frac{P_{i|m}P_{j|m}P_m}{P_j}\right]\beta_k X_{ik}\quad \text{for } i, j \text{ in same nest}$$

OGEV cross-elasticity = $-P_i\beta_k X_{ik}$ for i, j sharing no nest in common

Note that the OGEV cross-elasticity collapses to the MNL cross-elasticity equation for those alternatives that are separated by more than T time periods. The MNL proportional substitution property (or IIN) applies to those alternatives that do not share a nest in common and for which their covariance term is zero.

Distinct from the PCL model, it is no longer possible to express covariance terms (and associated correlation) in closed-form. Conceptually, this can be visualized by *attempting* to express the variance-covariance matrix for the OGEV model using the approach described for the PCL model. That is, the upper triangular of the matrix defines the covariance between alternatives i and j in nest ij that is "weighted" by how much of alternative i is allocated to nest ij whereas the lower triangular of the matrix defines the covariance between alternatives i and j in nest ij that is "weighted" by how much of alternative j is allocated to nest ij. Using these definitions, the following "variance-covariance" matrix would be defined as follows:

$$\Omega^* = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 1 & \tau_1(1-\mu_2^2) & 0 & 0 & 0 & 0 \\ \tau_0(1-\mu_2^2) & 1 & \tau_1(1-\mu_3^2) & 0 & 0 & 0 \\ 0 & \tau_0(1-\mu_3^2) & 1 & \tau_1(1-\mu_4^2) & 0 & 0 \\ 0 & 0 & \tau_0(1-\mu_4^2) & 1 & \tau_1(1-\mu_5^2) & 0 \\ 0 & 0 & 0 & \tau_0(1-\mu_5^2) & 1 & \tau_1(1-\mu_6^2) \\ 0 & 0 & 0 & 0 & \tau_0(1-\mu_6^2) & 1 \end{bmatrix} \end{array} \times \frac{\pi^2}{6\gamma^2}$$

However, all variance-covariance matrices *must be symmetric*, and so Ω^* collapses to the variance-covariance matrix for the OGEV model shown in Figure 4.4 only when a restrictive condition is applied, namely when $\tau_0 = \tau_1$. To emphasize, calculation of covariance terms (and associated correlations between alternatives) is no longer as straightforward as it was for the MNL, NL, and PCL models.

It was only recently that researchers have been able to derive the exact formula for utility correlations associated with two-level GEV models (such as the OGEV model). Specifically, Abbe, Bierlaire, and Toledo (2007) show that correlation between two alternatives is given as:

$$\mathrm{Corr}\left(U_i, U_j\right) = \mathrm{Corr}\left(\max_m\left(\ln\tau_{im} + \varepsilon_{im}\right), \max_m\left(\ln\tau_{jm} + \varepsilon_{jm}\right)\right) \tag{4.1}$$

$$\operatorname{Corr}\left(\varepsilon_{im},\varepsilon_{jm}\right)=\left(1-\mu_m^2\right)\delta_m\left(i,j\right) \tag{4.2}$$

where:

M	is the number of nests in a two-level GEV model,
τ_{im}	is the proportion of alternative i allocated to nest m, $\tau_{im} \geq 0\ \forall i,m$ and $\sum_{m=1}^{M} \tau_{im} = 1 \forall i$,
μ_m	is the logsum parameter associated with nest m, $0 < \mu_m \leq 1$,
$\delta_m(i,j)$	is a dummy variable equal to 1 if i and j are in nest m and 0 otherwise.

Equation (4.1) defines the total correlation between the alternatives i and j (remember that the true or theoretical utility U_i is composed of a systematic component of utility and an unobserved or stochastic error component, $U_i = V_i + \varepsilon_i$). Equation (4.2) defines the correlation between the error components for alternatives i and j and is the familiar expression seen in the context of the NL model.[1] As noted by Abbe, Bierlaire, and Toledo (2007), the relation between the overall correlation (Equation 4.1) and the underlying NL correlations (Equation 4.2) is made via a maximum operator. The overall correlation can be computed numerically from the joint cdf of the utilities:

$$\operatorname{Corr}\left(U_i,U_j\right)=\frac{6}{\pi^2}\int\int_{\mathbb{R}^2} x_i x_j \partial^2_{x_i x_j} F_{\varepsilon_i,\varepsilon_j}\left(x_i,x_j\right)\mathrm{d}x_i\mathrm{d}x_j-\frac{6\zeta}{\pi^2};\ \zeta = \text{Euler's Constant} = 0.577$$

where

$$F_{\varepsilon_i,\varepsilon_j}\left(x_i,x_j\right)=\exp\left(-\sum_{m=1}^{M}\left(\left(\tau_{im}e^{-x_i}\right)^{1/\mu_m}+\left(\tau_{jm}e^{-x_j}\right)^{1/\mu_m}\right)^{\mu_m}\right)$$

Clearly, calculations for correlations become computationally much more difficult as the underlying GEV model becomes more complex, despite the fact that closed-form expressions can still be obtained for choice probabilities. See Abbe, Bierlaire, and Toledo (2007) for further details, including implementation suggestions for numerically solving the resulting system of nonlinear equations.

As a final note with respect to the OGEV model, it is important to point out that the OGEV model can be extended to more than one adjacent time period, as shown in Figure 4.5. The OGEV probability, direct-elasticity, and cross-elasticity formulas discussed above are general and apply to OGEV models with more than one adjacent time period. The two-period OGEV model exhibits greater-than-MNL competition for itineraries that depart in the two time periods immediately before or immediately after. Further, itineraries one adjacent period away compete more than itineraries two adjacent periods away. For example, itineraries departing

1 The formulas in the original Abbe, Bierlaire, and Toledo (2007) work have been adapted to correspond to the definitions for logsum coefficients used in this text. It is also assumed that the logsum of the root node has been normalized to one.

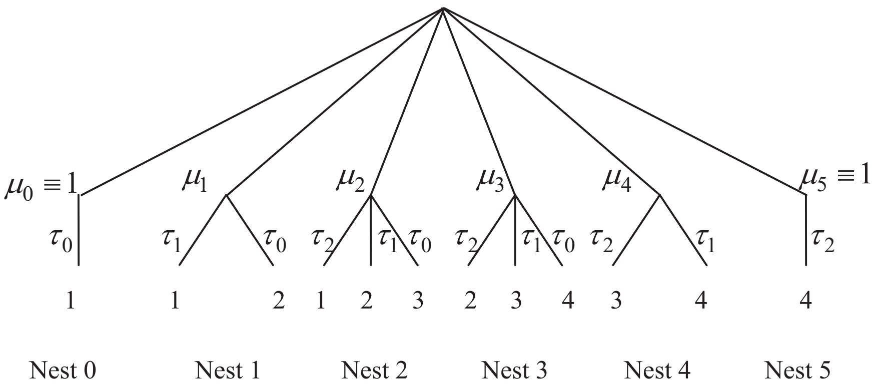


Figure 4.5 Ordered GEV model with two adjacent time periods

in period two compete more with itineraries departing in period three than with itineraries departing in period four.

In practice, the number of time-periods (or T) is determined by comparing model fits between a model that incorporates t time steps with a model that incorporates $t + 1$ time steps, e.g., the analyst compares the model fits between a one-step OGEV and a two-step OGEV, a two-step OGEV and a three-step OGEV, etc. Sufficient time periods have been incorporated when there is "little" improvement observed in the fit between two models. This formal test, called the non-nested hypothesis test, is described in detail in Chapter 7. Finally, it is important to note that the use of discrete time periods to allocate alternatives may result in awkward interpretations. For example, define three morning time periods as 8:01-9:00, 9:01-10:00, and 10:01-11:00. In an one-time period OGEV model, a flight departing at 9:59 would be expected to compete more with a flight departing at 9:01 (that belongs to the same nest) than with a flight departing at 10:01 (that belongs to an adjacent nest). However, from a practical perspective, the OGEV model offers more robust prediction results than MNL and NL models, the latter of which are more commonly encountered in airline itinerary choice applications.

Generalized Nested Logit

The MNL, NL, PCL and OGEV models are special cases of the generalized nested logit (GNL) model (Wen and Koppelman 2001). The GNL is more "general" in the sense that its nesting structures are not restrictive and both allocation and logsum parameters are estimated. An example of a GNL model is shown in Figure 4.6 for two train alternatives (one economy and one premium) and two air alternatives (one economy and one premium). The GNL model in Figure 4.6 contains four nests. The first and second nests incorporate increased competition among the train and air alternatives, respectively; the third and fourth nests incorporate increased competition among the economy and premium alternatives, respectively. In contrast to earlier

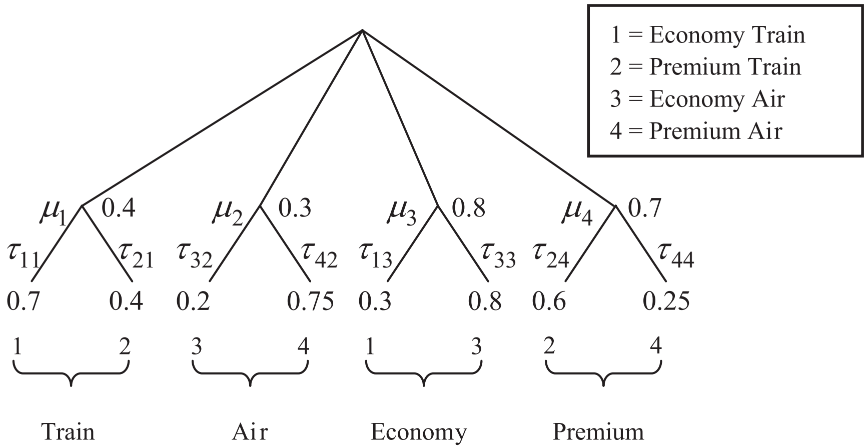


Figure 4.6 Generalized nested logit model

figures, parameters for logsum and allocation are shown in the figure, which is used to provide a numerical example of calculating GNL probabilities.

The GNL probability is given as:

$$P_i = \sum_m P_{i|m} \times P_m$$

$$= \sum_m \left[\frac{\left(\tau_{im} e^{V_i}\right)^{\frac{1}{\mu_m}}}{\sum_{j \in A_m} \left(\tau_{jm} e^{V_j}\right)^{\frac{1}{\mu_m}}} \times \frac{\left(\sum_{j \in A_m} \left(\tau_{jm} e^{V_j}\right)^{\frac{1}{\mu_m}} \right)^{\mu_m}}{\sum_l \left(\sum_{j \in A_l} \left(\tau_{jl} e^{V_j}\right)^{\frac{1}{\mu_l}} \right)^{\mu_l}} \right], \tag{4.3}$$

$$0 < \mu_m, \mu_l \leq 1, \sum_m \tau_{im} = 1 \; \forall \, i$$

where:

$j \in A_m$ is the set of all alternatives that belong to nest m,

m, l are indices used to sum over all nests,

τ_{im} are unknown allocation parameters that characterize the portion of alternative i assigned to a nest. Allocation parameters are non-negative, i.e., $\tau_{im} \leq 0$ and must sum to one for every alternative, which is equivalent to $\sum_m \tau_{im} = 1, \forall i \cdot$

μ_m is the logsum coefficient associated with nest m, $0 < \mu_m \leq 1, \forall\, m$.

Note that in the GNL model, both allocation and logsum parameters are estimated and alternatives may be allocated to one or more nests (the only condition is that the allocations associated with alternative *i* across all nests sum to one). For example, in Figure 4.6, 70 percent of alternative one (representing economy train) is allocated to the nest one (the train nest) and 30 percent is allocated to nest three (the economy nest), or $\tau_{11} + \tau_{13} = 1$. Similarly, 40 percent of alternative two (representing premium train) is allocated to nest one (the train nest) and 60 percent is allocated to nest four (the premium nest), or $\tau_{21} + \tau_{24} = 1$.

From an interpretation perspective, both allocation and logsum parameters provide information on the amount of competition among alternatives. In Figure 4.6, those nests with the smaller logsum parameters exhibit greater substitution, e.g., economy and premium train classes will compete more with each other ($\mu_1 = 0.4$) than economy train and economy air ($\mu_3 = 0.8$). Conceptually, larger allocations also lead to higher substitutions, e.g., 75 percent of alternative four, representing premium air, is allocated to the air competition nest, implying higher substitution between premium air and economy air ($\tau_{42} = 0.75$) than premium air and premium train ($\tau_{44} = 0.25$). Of course, as seen in Equations (4.2) and (4.3), the exact value for the amount of competition or substitution between two alternatives is ultimately given as a function of allocation and logsum parameters, which may not be straightforward to calculate.

To visualize the calculations of GNL probabilities, assume an individual's decision of whether to take economy train, premium train, economy air, or premium air is expressed as a function of time and cost (as well as an intercept term):

$$V_i^n = \alpha_i + \beta_1\left(Time_i^n\right) + \beta_2\left(Cost_i^n\right)$$

Suppressing the index representing individual *n* for notational convenience, assume the utility function for the four alternatives (faced with specific time and costs) is given as:

$$V_1 = 1 - 0.075(5\text{ hrs}) - 0.0015(\$300) = 0.175$$

$$V_2 = 0.5 - 0.075(5\text{ hrs}) - 0.0015(\$400) = -0.475$$

$$V_3 = 2.5 - 0.075(3\text{ hrs}) - 0.0015(\$350) = 1.75$$

$$V_4 = 0 - 0.075(3\text{ hrs}) - 0.0015(\$750) = -1.275$$

The calculation of probabilities for each alternative can be visualized by revisiting Equation (4.3) as being composed of four terms, as illustrated below:

$$P_i = \sum_m \left[\frac{\overbrace{\left(\tau_{im} e^{V_i}\right)^{\frac{1}{\mu_m}}}^{\text{A TERM}}}{\underbrace{\sum_{j \in A_m} \left(\tau_{jm} e^{V_j}\right)^{\frac{1}{\mu_m}}}_{\text{B TERM}}} \times \frac{\overbrace{\left(\sum_{j \in A_m} \left(\tau_{jm} e^{V_j}\right)^{\frac{1}{\mu_m}}\right)^{\mu_m}}^{\text{C TERM}}}{\underbrace{\sum_l \left(\sum_{j \in A_l} \left(\tau_{jl} e^{V_j}\right)^{\frac{1}{\mu_l}}\right)^{\mu_l}}_{\text{D TERM}}} \right]$$

where:

A TERM is computed for each alternative in nest m,
B TERM is the sum of all A TERMS in nest m,
C TERM is the B TERM raised to the logsum coefficient for nest m, or μ_m,
D TERM is the sum, overall all nests, of C TERMS.

Table 4.2 contains intermediate calculations used to compute the probabilities for each alternative. The analyst begins by calculating A TERMS. Because there are two alternatives associated with the first nest, there are two A TERMS given as:

$$\text{A TERM for alternative one in nest one} = \left(\tau_{im} e^{V_i}\right)^{\frac{1}{\mu_m}} = \left(0.7e^{0.175}\right)^{\frac{1}{0.4}} = 0.6350$$

$$\text{A TERM for alternative two in nest one} = \left(\tau_{im} e^{V_i}\right)^{\frac{1}{\mu_m}} = \left(0.4e^{-0.475}\right)^{\frac{1}{0.4}} = 0.0309$$

The B TERM associated with nest one is simply the sum of these two A TERMS, or 0.6350 + 0.0309 = 0.6658. The C TERM associated with nest one is also straightforward, i.e., the B TERM raised to the logsum coefficient for nest one, or:

$$\text{C TERM for nest one} = \left(0.6658\right)^{0.4} = 0.8499$$

The process is repeated for each of the nests, after which the D TERM is calculated as the sum of all of the C TERMS, or 0.8499 + 1.1521 + 4.7540 + 0.3967 = 7.153. With the intermediate calculations of Table 4.2 completed, the probability of alternative one is given as the sum of probabilities calculated for nests one and three (the nests that alternative one belongs to), or:

$$P_1 = \frac{A_1}{B_1} \times \frac{C_1}{D} + \frac{A_3}{B_3} \times \frac{C_3}{D}$$

$$= \frac{0.6350}{0.6658} \times \frac{0.8499}{7.1526} + \frac{0.2763}{7.0198} \times \frac{4.7540}{7.1526} = 0.1395.$$

Similar calculations apply for alternatives two, three, and four.

Table 4.2 Intermediate calculations for GNL probabilities

	A TERM				B TERM	C TERM	D TERM
	ALT 1	ALT 2	ALT 3	ALT 4			
Nest 1	0.6350	0.0309	0	0	0.6658	0.8499	7.153
Nest 2	0	0	1.5977	1.6031	1.6031	1.1521	7.153
Nest 3	0.2763	0	6.7434	0	7.0198	4.7540	7.153
Nest 4	0	0.2446	0	0.0223	0.2669	0.3967	7.153
Probability	0.1395	0.0563	0.7990	0.0052			

The elasticity and cross-elasticity formulas for the GNL are similar to those for the OGEV model, except the summation applies across all nests (not just those in adjacent time periods). Also, note that if two alternatives do not share any nest in common, the cross-elasticity equation collapses to that for the MNL model:

$$\text{GNL direct-elasticity} = \left[(1-P_i) + \sum_m \left(\frac{1-\mu_m}{\mu_m} \right) \frac{P_{i|m} P_m \left(1-P_{i|m}\right)}{P_i} \right] \beta_k X_{ik}$$

$$\text{GNL cross-elasticity} = -\left[P_i + \sum_m \left(\frac{1-\mu_m}{\mu_m} \right) \frac{P_{i|m} P_{j|m} P_m}{P_j} \right] \beta_k X_{ik}$$

The CNL (Vovsha 1997) and the Gen-MNL (Swait 2000) models shown in Table 4.1 are constrained versions of the GNL. Specifically, the CNL is equivalent to a GNL, except the CNL constrains all logsums to be equal. Similarly, the Gen-MNL is equivalent to a GNL, except the Gen-MNL constrains all allocation parameters

to equal one. These constraints can be observed by comparing formulas for probabilities, direct-elasticities, and cross-elasticities formulas associated with the different models. The GNL is also a relaxed version of models discussed in the next section that include the product differentiation (PD) model (Bresnahan Stern and Trajtenberg 1997), and the weighted nested logit (WNL) model.

There are several important points of interest related to the GNL model. First, given that the GNL is one of the most general GEV-formulations, it has often used in empirical applications or as a "baseline" against which alternative discrete choice models (most recently, those based on a mixed logit formulation) are compared (e.g., see Chiou and Walker 2007; Gopinath Schofield Walker and Ben-Akiva 2005; Hess Bierlaire and Polak 2005a; Munizaga and Alvarez-Daziano 2001, 2002).

Researchers in Europe generalized Vovsha's CNL model parallel to the development of the GNL model by Wen and Koppelman. Thus, the terms "generalized nested logit" and "cross-nested logit" are often used interchangeably in the literature to refer to the model shown in Figure 4.6 and defined by Equation 4.3. In this text, GNL will be used to refer to this model to more clearly distinguish between the original CNL model proposed by Vovsha in 1997 that constrains all logsums to be equal and the GNL model proposed by Wen and Koppelman in 2001 that estimates both logsum and allocation parameters.

"Weighted" Combinations of GEV Models Used for Itinerary Choice Applications

From an applications perspective, the OGEV, GNL, and other advanced GEV models are particularly useful for capturing competitive dynamics among airline itineraries. This is because competition across itineraries occurs in multiple dimensions, including time of day (which has an inherent ordering), carrier, level of service (non-stop, direct, connection), and fare class (first/business, unrestricted high yield, restricted low yield, etc.). In the early 2000s, Coldren and Koppelman compared model fits across several GEV-based models presented thus far (Coldren 2005; Coldren and Koppelman 2005a). These include the MNL, two-level NL, three-level NL, OGEV, and two-level GNL models. Based on the fact that their models were detecting multiple dimensions of competition, they developed several new GEV models that contain three levels. Three of the models they proposed, namely the weighted nested logit (WNL), nested-weighted nested logit (N-WNL) and ordered GEV-nested logit[2] (OGEV-NL) are discussed in the following sections.

---

2 This model was originally called the NL-OGEV model, but will be renamed as the OGEV-NL model in this text in order to provide a consistent naming convention to represent the "upper level structure—lower level structure."

Product Differentiation (PD) and Weighted NL (WNL) Models

The motivation for the weighted nested logit (WNL) model for itinerary choice applications can be seen in Figure 4.7. That is, competition among airline itineraries occurs across multiple dimensions. The nest on the left-hand side of Figure 4.7 captures increased substitution across brands (i.e., the addition of an American Airlines flight to a network schedule is expected to compete more with existing American Airlines flights than those operated by other carriers). Similarly, the nest on the right-hand side of Figure 4.7 captures increased substitution across time of day (i.e., the addition of an American Airlines flight in time period two is expected to compete more with flights that depart in time period two than those that depart in periods one or three). In this sense, the WNL model can be viewed as an extension of the two-level NL model in that increased competition is incorporated across both the carrier and time of day dimensions.

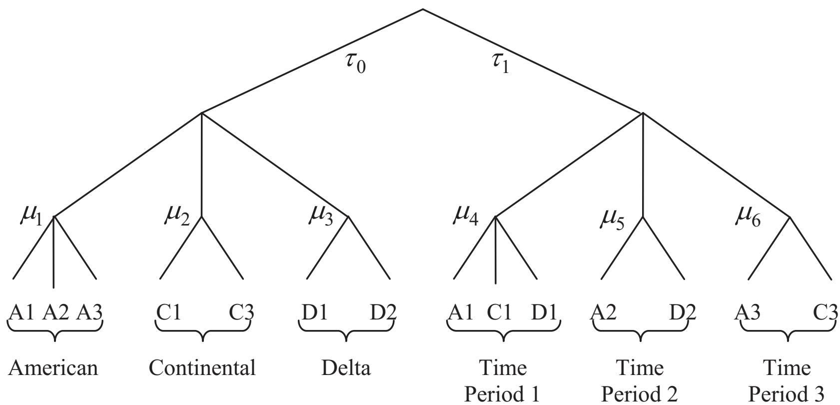


Figure 4.7 "Weighted" nested logit model

The WNL model is equivalent to the product differentiation (PD) model proposed in 1997 by Bresnahan, Stern, and Trajtenberg. However, the underlying motivations for the models are slightly different. Whereas the WNL arises from the recognition that substitution patterns are formed by "weighting" underlying "NL models," the PD model arises from the recognition that products can be clustered into separate groups based on one or more product dimensions; those products in the same cluster are expected to compete more with each other than with those products in other clusters.

Formally, the WNL probability is given as:

$$P_i = \sum_{f \in F} P_{i|f} \times P_f$$

$$= \sum_{f \in F} \left[\frac{e^{\frac{V_i}{\mu_f}}}{\sum_{k \in A_f} e^{\frac{V_k}{\mu_f}}} \times \frac{w_f \left(\sum_{k \in A_f} e^{\frac{V_k}{\mu_f}} \right)^{\mu_f}}{\sum_{c \in F} w_c \left(\sum_{s \in A_c} e^{\frac{V_s}{\mu_c}} \right)^{\mu_c}} \right] \tag{4.4}$$

$$0 < \mu_f \leq 1, \sum_{c \in F} w_c = 1$$

where:

F is the product dimension set,

$k \in A_f$ is the set of all alternatives (products) that belong to the same cluster characterized by dimension f,

w_f is the weight for dimension f. Weight parameters are non-negative, i.e., $w_f \geq 0$ and must sum to one, which is equivalent to

$$\sum_{f \in F} w_f = 1$$

μ_f is the logsum coefficient associated with all clusters (nest) along dimension f, $0 < \mu_f \leq 1$, $\forall f$.

The variance-covariance matrix for the WNL model shown in Figure 4.7 is given as:

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{bmatrix} 1 & \tau_0\left(1-\mu_1^2\right) & \tau_0\left(1-\mu_1^2\right) & \tau_1\left(1-\mu_4^2\right) & 0 & \tau_1\left(1-\mu_4^2\right) & 0 \\ \tau_0\left(1-\mu_1^2\right) & 1 & \tau_0\left(1-\mu_1^2\right) & 0 & 0 & 0 & \tau_1\left(1-\mu_5^2\right) \\ \tau_0\left(1-\mu_1^2\right) & \tau_0\left(1-\mu_1^2\right) & 1 & 0 & \tau_1\left(1-\mu_6^2\right) & 0 & 0 \\ \tau_1\left(1-\mu_4^2\right) & 0 & 0 & 1 & \tau_0\left(1-\mu_2^2\right) & \tau_1\left(1-\mu_4^2\right) & 0 \\ 0 & 0 & \tau_1\left(1-\mu_6^2\right) & \tau_0\left(1-\mu_2^2\right) & 1 & 0 & 0 \\ \tau_1\left(1-\mu_4^2\right) & 0 & 0 & \tau_1\left(1-\mu_4^2\right) & 0 & 1 & \tau_0\left(1-\mu_3^2\right) \\ 0 & \tau_1\left(1-\mu_5^2\right) & 0 & 0 & 0 & \tau_0\left(1-\mu_3^2\right) & 1 \end{bmatrix} \end{array} \times \frac{\pi^2}{6\gamma^2} \tag{4.5}$$

To more clearly visualize how the WNL model can be transformed into a constrained version of the GNL model, the w weights associated with the carrier and time of day nests reflected in Equation 4.4 have been assigned values of τ_0 and

τ_1, respectively, in the variance-covariance matrix. In the PD model representation, the carrier and time-periods would be represented as two product dimensions. In the WNL model representation, the carrier and time periods would be represented as the weighted combination of two variance-covariance matrices, one that represents the amount of competition along the carrier "dimension" and the other that represents the amount of competition among the "time period" dimension. Importantly, in this particular model, it is possible to directly compute covariance and correlation terms and interpret the WNL model as a weighted combination of two variance-covariance matrices due to the specific way in which the model was defined. Formally, the carrier substitution (which receives a weight of τ_0) is given by the symmetric matrix:

$$\begin{array}{c c}
 & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} &
\begin{bmatrix}
1 & 1-\mu_1^2 & 1-\mu_1^2 & 0 & 0 & 0 & 0 \\
1-\mu_1^2 & 1 & 1-\mu_1^2 & 0 & 0 & 0 & 0 \\
1-\mu_1^2 & 1-\mu_1^2 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1-\mu_2^2 & 0 & 0 \\
0 & 0 & 0 & 1-\mu_2^2 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1-\mu_3^2 \\
0 & 0 & 0 & 0 & 0 & 1-\mu_3^2 & 1
\end{bmatrix} \times \dfrac{\tau_0 \pi^2}{6\gamma^2}
\end{array}$$

whereas the time of day substitution (which receives a weight of τ_1) is given by the symmetric matrix:

$$\begin{array}{c c}
 & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} &
\begin{bmatrix}
1 & 0 & 0 & 1-\mu_4^2 & 0 & 1-\mu_4^2 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1-\mu_5^2 \\
0 & 0 & 1 & 0 & 1-\mu_6^2 & 0 & 0 \\
1-\mu_4^2 & 0 & 0 & 1 & 0 & 1-\mu_4^2 & 0 \\
0 & 0 & 1-\mu_6^2 & 0 & 1 & 0 & 0 \\
1-\mu_4^2 & 0 & 0 & 1-\mu_4^2 & 0 & 1 & 0 \\
0 & 1-\mu_5^2 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} \times \dfrac{\tau_1 \pi^2}{6\gamma^2}
\end{array}$$

Note that by decomposing the variance-covariance matrices along their "product" dimensions, several important characteristics inherent in the PD and WNL models become clear. First, each alternative must appear exactly once in for each dimension, i.e., each itinerary appears once in the carrier nest and once in the time of day nest. This structure also results in a symmetric substitution pattern for each pair of alternatives, as seen in the variance-covariance matrix.

Second, it is important to note that the combined variance-covariance matrix shown in Equation 4.5 is equivalent to the GNL model shown in Figure 4.8. Thus, there are three interpretations that can be associated with the weights. In the PD model, the weights can be thought of as the relative importance of the carrier and time of day product characteristics in determining substitution patterns. A larger weight associated with the carrier product dimension implies larger recapture rates within a given carrier, whereas a larger weight associated with the time of day dimension implies customers are more likely to travel on competing brands that depart closer to the customers' preferred times. A similar interpretation for the weights applies for the WNL model. However, in the WNL model, the weights can also be interpreted in terms of weighted combinations of variance-covariance structures that are common to the literature (in this case, the weighted combination of two NL models). In the GEV model, the weights can be interpreted as allocation parameters, or the percentage of each alternative that is assigned to respective carrier and time of day nests.

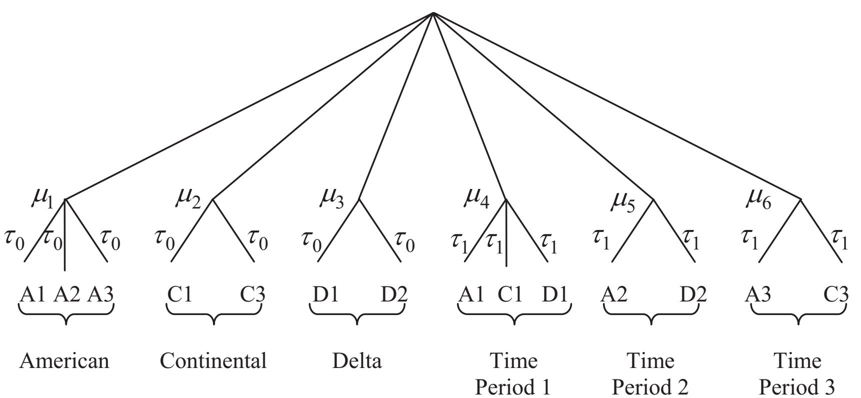


Figure 4.8 GNL representation of weighted nested logit model

Finally, and most critically, it is important to note that although different interpretations for the weights arise from the PD, WNL, and GNL representation, the probability of selecting an alternative is identical across the three models. That

is, in the WNL model,[3] it is not possible to express the probability of selecting alternative *i* as:

$$P_i = \quad w_{carrier}\left(P_{i|carrier} \times P_{carrier}\right) + w_{time\ of\ day}\left(P_{i|time\ of\ day} \times P_{time\ of\ day}\right)$$

The weights must be applied at the lower-level, or as allocation parameters, as in the GNL model representation. This is required in order to ensure that the marginal probabilities associated with choosing the carrier nest or time of day nest sum to one.

Given that the PD and WNL models represent special cases of the GNL model, their direct- and cross-elasticities are equivalent. Using set notation to represent product dimensions (or combinations of multiple NL nests), the direct- and cross-elasticities for the PD and WNL model is given, respectively, as:

PD and WNL direct-elasticity: $\left[(1-P_i)+\sum_{f\in F}\left(\frac{1-\mu_f}{\mu_f}\right)\frac{P_{i|f}(1-P_{i|f})P_f}{P_i}\right]\beta_k X_{ik}$

PD and WNL cross-elasticity: $-\left[P_i+\sum_{f\in F}\left(\frac{1-\mu_f}{\mu_f}\right)\frac{P_{i|f}P_{j|f}P_f}{P_j}\right]\beta_k X_{ik}$

Nested-Weighted Nested Logit

The WNL model can be interpreted as a weighting of one or more two-level NL models. However, there are three-level structures representing other combinations of GEV models that belong to the more general NetGEV family that may also be appropriate in the context of itinerary choice models. For example, the nested weighted nested logit (N-WNL) model, shown in Figure 4.9, groups (or nests) itineraries according to departure time periods as upper-level nests and then combines two NL models to represent carrier and itinerary level of service (non-stop, connecting) competition structures at the lower level. From an interpretation perspective, alternatives that share a nest lowest in the tree compete most with each other. Among the lower-level nests, the nest with the smallest value of μ_{fm} exhibits the largest competition among alternatives in the nest. For example, a value of $\mu_{11} < \mu_{12}$ implies that American Airline itineraries that depart in time period one compete more with each other (i.e., exhibit stronger brand loyalty) than flights operated by Delta Air Lines that depart in time period one. By grouping itineraries at the higher level by departure time periods, the N-WNL model allows the mix of carrier and level of service competition to vary by time period. However, this grouping also imposes the assumption that itineraries that depart in different time

3 Note that this has been modified from the Coldren and Koppelman (2005a) paper.

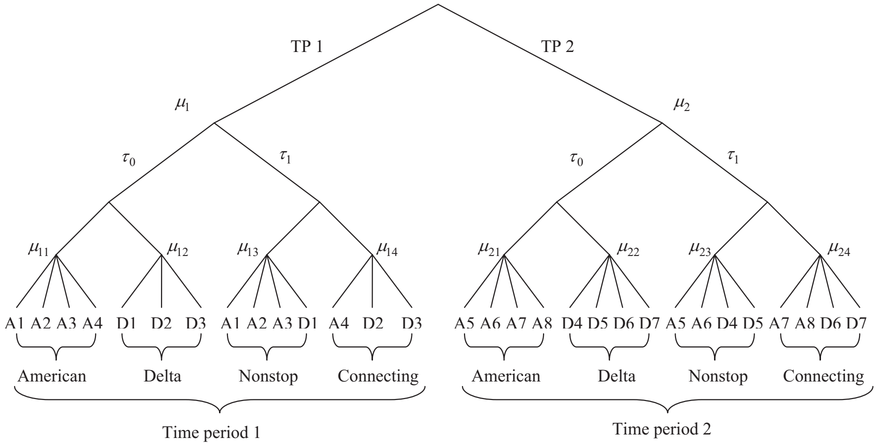


Figure 4.9 Nested-weighted nested logit model

periods exhibit the IIA property (or proportional substitution property inherent in the MNL model).

The N-WNL probability is given as:

$$\frac{e^{\mu_m \Gamma_m}}{\sum_{l=1}^{M} e^{\mu_l \Gamma_l}} \times \sum_{e \in E} \left[\frac{e^{\frac{V_i}{\mu_{fm}}}}{\sum_{k \in A_f} e^{\frac{V_k}{\mu_{fm}}}} \times \frac{w_f e^{\frac{\mu_{fm}}{\mu_m} \Gamma_{em}}}{\sum_{c \in F} w_c e^{\frac{\mu_{cm}}{\mu_m} \Gamma_{cm}}} \right],$$

$$0 < \mu_{fm} \le \mu_m \le 1, \sum_{f \in F} w_f = 1, \quad \Gamma_{fm} = \ln\left(\sum_{k \in A_f} e^{\frac{V_k}{\mu_{fm}}} \right), \quad \Gamma_m = \ln\left(\sum_{c \in F} w_c e^{\frac{\mu_{cm}}{\mu_m} \Gamma_{cm}} \right)$$

where:

- M is the number of upper-level nests,
- E is the product dimension set of the lower-level tree,
- $k \in A_f$ is the set of all alternatives (products) that belong to the same cluster characterized by dimension f of the lower-level tree,
- w_f is the weight for (lower-level) dimension f. Weight parameters are non-negative, i.e., $w_f \ge 0$ and must sum to one, which is equivalent to

$$\sum_{f \in F} w_f = 1$$

μ_m is the logsum coefficient associated with nest m of the upper-level tree, f, $0 < \mu_m \leq 1, \forall\, m$,

μ_{fm} is the logsum coefficient that characterizes the portion of the lower-level dimension f assigned to a (upper level) nest m, $0 < \mu_{fm} \leq \mu_m \leq 1, \forall\, m, f$.

Note that the first product of the N-WNL probability represents the upper level of the tree and has the same structure as the marginal probability of selecting nest m in the NL model. The terms that are summed over $f \in F$ represent the lower level of the tree and have identical structure to the WNL model. In order to ensure that covariance terms are non-negative, logsums associated with lower nests, μ_{fm}, must be smaller than logsums associated with their higher-level nests, μ_m.

The variance-covariance matrices, direct-elasticities and cross-elasticities for the hybrid N-WNL model become more complex than those for the two-level GNL models discussed thus far (see Table 4.4 at the end of this chapter for the direct- and cross-elasticity formulas). Similar to the OGEV model, the N-WNL covariance terms can no longer be expressed in closed-form. Further, unlike the very unique structure of the W-NL model, the N-WNL cannot be interpreted as a pure "weighting" of different variance-covariance matrices, one constructed from the "nested" subcomponent and the second constructed from the "weighted nested logit" subcomponent.

From an interpretation perspective, alternatives that share a nest lowest in the tree compete most with each other. For alternatives departing in the same time period, those that share two lower nests in common exhibit higher competition than those that share only one lower nest in common. Similarly, those that share one lower nest in common exhibit higher competition than those that share no lower nests in common. By grouping alternatives at the highest level of the nest by time of day, itineraries that depart in different time periods have zero covariance and exhibit the IIA property (or proportional substitution property). The OGEV-NL model, presented in the next section, is an alternative hybrid GEV model that was designed to address this limitation.

Ordered GEV—Nested Logit

The Ordered GEV—Nested Logit model shown in Figure 4.10 is similar to the N-WNL model in the sense that it is a hybrid, or mixture, of two GEV models. The OGEV-NL model combines a two-step OGEV model as the upper-level structure with a NL carrier competition nest at the lower level. Note that although the original OGEV-NL proposed by Coldren and Koppelman (2005a) constrained logsums to be equal at the OGEV and NL levels, theoretically it is possible to estimate separate logsum parameters for each nest. From an interpretation perspective, alternatives that share a common carrier and departure time period will compete most each other. In contrast to the W-NL model, alternatives that are one departure time period apart will compete more with each other whereas alternatives that are separate by two or more departure time periods will exhibit the proportional substitution (or IIA) property. The use of an OGEV departure time structure for the

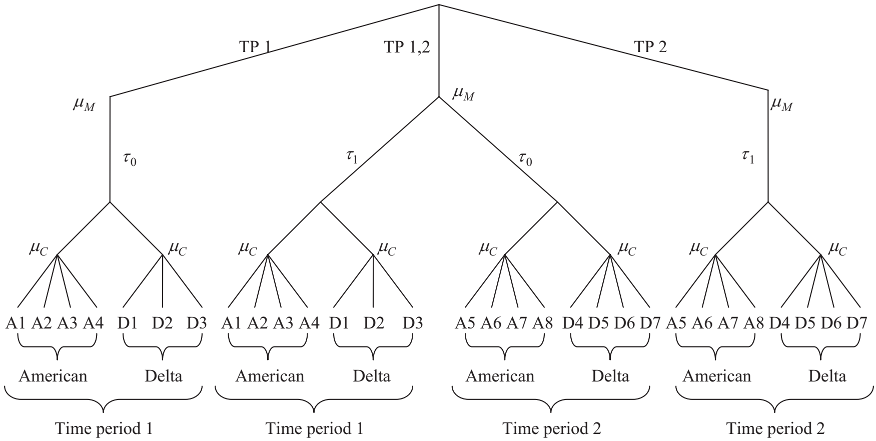


Figure 4.10 OGEV-NL model

upper-level nest will also result in covariance and correlation terms that are much more difficult to quantify.

The probability for the OGEV-NL model is given as:

$$\sum_{m=1}^{i+T}\left[\frac{e^{\mu_M\Gamma_m}}{\sum_{r=1}^{J+T} e^{\mu_M\Gamma_r}}\times\frac{e^{\frac{\mu_C}{\mu_M}\Gamma_{mc}}}{\sum_{d\in A_m} e^{\frac{\mu_C}{\mu_M}\Gamma_{md}}}\times\frac{\tau_{m-i}e^{\frac{V_i}{\mu_C}}}{\sum_{j\in A_{m,c}}\tau_{m-j}e^{\frac{V_j}{\mu_C}}}\right],\quad 0<\mu_C\le\mu_M\le 1,\ \sum_{m=1}^{T}\tau_m=1,$$

$$\Gamma_{mc}=\ln\left(\sum_{j\in A_{m,c}}\tau_{m-j}e^{\frac{V_j}{\mu_C}}\right),\quad \Gamma_m=\ln\left(\sum_{c\in A_m} e^{\frac{\mu_C}{\mu_M}\Gamma_{mc}}\right)$$

where:

T is the number of adjacent time periods in the upper-level OGEV model,

J is the total number of alternatives in the upper-level OGEV model,

$d\in A_m$ is the set of all (lower-level NL) clusters that belong to nest alternatives that belong to nest *m* in the upper-level OGEV model,

$j\in A_{m,c}$ is the set of all alternatives that belong to the nest of lower NL nests *c* in upper-OGEV nest *m*,

$\tau_{m\text{-}i}$ are unknown allocation parameters that characterize the portion of alternative *i* assigned to a nest. Allocation parameters are non-negative,

i.e., $\tau_{m\text{-}i} \geq 0$ and must sum to one for every alternative, which is equivalent to

$$\sum_{m=1}^{J+T} \tau_m = 1$$

μ_M is the logsum coefficient associated with the upper OGEV nest,
μ_C is the logsum coefficient associated with the lower NL nest.

The first product represents the lower-level NL structure, whereas the second two products represent the upper-level OGEV structure. The direct- and cross-elasticities are given, respectively, as:

$$\text{OGEV-NL direct-elasticity} = \left[\left(\frac{1}{\mu_C} - P_i \right) + \frac{1}{P_i} \sum_{m=i}^{i+T} \left(\begin{array}{l} \left(1 - \frac{1}{\mu_M} \right) P_m P_{c|m}^2 P_{i|c,m}^2 + \\ \left(\frac{1}{\mu_M} - \frac{1}{\mu_C} \right) P_m P_{c|m} P_{i|m,c}^2 \end{array} \right) \right] \beta_k X_{ik}$$

$$\text{OGEV-NL cross-elasticity} = \left[-P_i + \frac{1}{P_j} \sum_{m=j}^{j+T} \left(\begin{array}{l} \left(1 - \frac{1}{\mu_M} \right) P_m P_{c|m}^2 P_{i|m,c} P_{j|m,c} + \\ \left(\frac{1}{\mu_M} + \frac{1}{\mu_C} \right) P_m P_{c|m} P_{i|m,c} P_{j|c,m} \end{array} \right) \right] \beta_k X_{ik}$$

Bringing it all Together—Which Model?

In light of the numerous choice models that are available, the question naturally arises: Which one should be used in practical applications? The response to this question clearly depends on the application, available data, and staff time and expertise available to formulate and estimate choice models. On the one hand, there is a close relationship between the observed variables that are included in the systematic portion of the utility function and the unobserved variables that are reflected in the error components. Using an overly-simplistic specification for the systematic portion of utility may result in the "need" to use more complex model structures. Thus, it is important to strike a balance between the amount of time an analyst spends calibrating a utility function (within a MNL framework) and the amount of time an analyst spends investigating more complex formulations. Similarly, it is important to recognize that as the complexity of the model increases, so too do programming requirements—both for the initial estimation phase and the post-implementation phase. That is, researchers who want to develop new choice models or use more recent choice models, such as the OGEV-NL model,

will need to write their own log likelihood functions[4] or use customized software such as BIOGEME (Bierlaire 2003, 2008) or ELM (Elm-Works Inc. 2008). These more complex probability expressions will also need to be programmed for any implementation. These two key points argue in favor of using more simplistic model structures, with a well-specified utility function.

On the other hand, it can also be argued that the airline industry is one of the industries that can dramatically benefit from the investigation of more complex model formulations. This is because even small gains in forecasting accuracy can translate into millions of dollars of additional revenue. Further, it is the authors' opinion that airline industry applications will continue to drive new discrete choice model developments. This is due to two key influences. First, given the numerous competitive dimensions associated with airline itineraries (e.g., carrier, itinerary level of service, fare class and associated product restrictions, departure times), the prediction of individuals' itinerary choices (or related booking class) will likely benefit by incorporating more flexible variance-covariance matrices. Second, the airline industry is particularly well positioned to leverage on-line data sources as they become available. This on-line data effectively capture choice sets available at the time of booking on the carrier of interest (as well as its competitors). As a side note, the volume of data that needs to be processed for airline industry applications (that are orders of magnitude larger than datasets from urban travel demand applications which drove the initial development of many of the early discrete choice models) will also force a reexamination of the optimization algorithms used to solve for the parameters of choice models, as well as the development of more automated processes that uncover the "best" model structures (in terms of the covariance components that fit the data the best and exhibit logical substitution patterns).

When selecting which choice model is most appropriate, it is also important to note that the models presented in this chapter, although representative, are not exhaustive. The GEV models highlighted in the context of itinerary choice applications are still restrictive in the sense that they cannot incorporate random taste variation, cannot incorporate correlation across observations (as is the case with panel data or data in which there are repeat observations by the same individual), and impose the assumption of homoscedastic variance. The mixed logit model, discussed in depth in Chapter 6, is an alternative model that can be used in problem contexts in which it is important to relax these assumptions. Theoretically, the mixed logit is particularly attractive in the sense that it can approximate any discrete choice model (Dalal and Klein 1988; McFadden and Train 2000). However, its choice probabilities can no longer be expressed in closed-

4 Analysts do not need to program their own optimization routines to solve for these parameters, i.e., many software programs, including Gauss (Aptech 2008), have standard optimization routines. These routines use two key pieces of information as inputs—the log likelihood equation and partial derivatives of log likelihood function with respect to the vector of parameters.

form and must be numerically evaluated, which may be an important consideration in aviation contexts that require daily processing of millions of observations.

Summary of Main Concepts

This chapter presented an overview of different discrete choice models, emphasizing those models that fall within the GEV class that allocate alternatives to more than one nest. Two other key concepts related to GEV models that researchers need in order to create their own models will be expanded upon in subsequent chapters. Specifically, Chapter 5 covers the theoretical requirements associated with GEV-generating functions and introduces the NetGEV model, which has proved particularly useful for developing normalizing rules required to ensure a new model is properly identified. Chapter 6 covers the mixed logit model.

The most important concepts covered in this chapter include the following:

- The development of discrete choice models has been a very active area of research. Dozens of models have been developed, several that were specifically motivated by airline applications.
- Models that belong to the GEV class can relax the IIA and/or IIN assumptions by including covariance terms that are created by allocating alternatives to two or more nests.
- The ability to obtain closed-form probability expressions for GEV models arises from the assumption that the total variance is identically distributed across alternatives.
- It is often useful to distinguish between GEV models that contain two levels (GNL models) and GEV models that contain three levels (NetGEV models). However, it should be recognized that GNL models are a special case of NetGEV models.
- Although it is fairly straightforward to derive closed-form expressions for GNL and NetGEV direct- and cross-elasticities, the calculation of covariance and correlation terms can be much more complex. For many GEV models, it is not possible to express correlations in closed-form; exact estimates of these correlations require solving for a nonlinear system of equations using numerical estimation methods.
- When developing new discrete choice models, it is important to include an assessment of the maximum amount of correlation that can be accommodated between two alternatives and to ensure that the proposed model is uniquely identified (or properly normalized).
- The PCL model, although simple, is useful for visualizing why there are maximum limits on the amount of correlation that can be accommodated between any pair of alternatives.
- The OGEV model is used in applications in which the ordering of alternatives has a physical meaning, e.g., to capture time of day competition effects among itineraries.

- The PD and WNL models are equivalent, but arise from different motivations. The WNL arises from the recognition that substitution patterns are formed by weighting underlying NL models. The PD model arises from the recognition that products can be clustered into separate groups based on one or more product dimensions.
- The WNL and PD models exhibit very specific nesting structures and allocation of alternatives to nests that results in the ability to express their associated variance-covariance matrices in closed-form and interpret this matrix as a "pure weighting" of two or more underlying NL variance-covariance matrices. In contrast, hybrid models such as the N-WNL and OGEV-NL model are often viewed as a "weighting" of different GEV models, but it is important to note that their resulting variance-covariance matrices do not necessarily lead to variance-covariance matrices that are easy to compute.
- In practice, it is common to impose constraints on the relationships among logsum and/or allocation parameters to avoid empirical identification issues.
- GEV models, including the MNL, NL, and GNL models, are commonly used in practice. However, these models are still limited in the sense they cannot incorporate random taste variation, correlation in errors across observations, and unequal variance. The mixed logit model, discussed in Chapter 6, is commonly used in applications in which it is important to relax these assumptions.

Table 4.3 Summary of probabilities for select GEV models

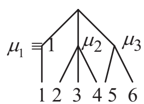
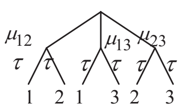
Model	Probability, P_i (expressed as $P_{i\|m} \times P_m$ where applicable)	Key Properties
Multinomial Logit (MNL) McFadden 1973 	$$\frac{e^{V_i}}{\sum_J e^{V_j}}$$	• Correlation = 0 • Independence of Irrelevant Alternatives (IIA) • No nests
Nested Logit (NL) McFadden 1978, Williams 1978 	$$\frac{e^{\frac{V_i}{\mu_m}}}{\sum_{j \in A_m} e^{\frac{V_j}{\mu_m}}} \times \frac{e^{V_m + \mu_m \Gamma_m}}{\sum_{l=1}^{M} e^{V_l + \mu_l \Gamma_l}},$$ $$\Gamma_m = \ln\left(\sum_{j \in A_m} e^{\frac{V_j}{\mu_m}}\right), 0 < \mu_m \le 1$$	• Correlation = $1 - \mu_m^2$ for *i,j* in nest *m*, 0 otherwise • Independence of Irrelevant Nests (IIN) • Alternatives allocated to exactly one nest (may be a degenerate nest with one alternative)
Paired Combinatorial Logit (PCL) Chu 1989, Koppelman & Wen 1998b 	$$\sum_{j \ne i} \left[\frac{\left(\tau e^{V_i}\right)^{\frac{1}{\mu_{ij}}}}{\left(\tau e^{V_i}\right)^{\frac{1}{\mu_{ij}}} + \left(\tau e^{V_j}\right)^{\frac{1}{\mu_{ij}}}} \times \frac{\left(\left(\tau e^{V_i}\right)^{\frac{1}{\mu_{ij}}} + \left(\tau e^{V_j}\right)^{\frac{1}{\mu_{ij}}}\right)^{\mu_{ij}}}{\sum_{r=1}^{J-1} \sum_{s=r+1}^{J} \left(\left(\tau e^{V_r}\right)^{\frac{1}{\mu_{rs}}} + \left(\tau e^{V_s}\right)^{\frac{1}{\mu_{rs}}}\right)^{\mu_{rs}}} \right],$$ $$0 < \mu_{ij} \le 1, \quad \tau = \frac{1}{N-1}$$	• Correlation = $1 - \mu_{ij}^2$ for *i,j* in nest *m*, 0 otherwise • Each pair of alternatives shares one nest together • Each alternative is allocated to $\tau = 1/(N-1)$ nests • All covariance terms can be non-negative, but max correlation between any pair of alternatives = τ

Table 4.3 *Continued*

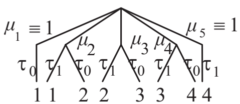
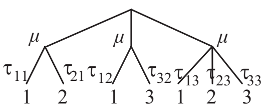
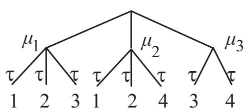
Ordered Generalized Extreme Value (OGEV) Small 1987 	$\sum_{m=i}^{i+T}\left[\frac{\left(\tau_{m-i}e^{V_i}\right)^{\frac{1}{\mu_m}}}{\sum_{j\in A_m}\left(\tau_{m-j}e^{V_j}\right)^{\frac{1}{\mu_m}}}\times\frac{\left(\sum_{j\in A_m}\tau_{m-j}e^{\frac{V_j}{\mu_m}}\right)^{\mu_m}}{\sum_{r=1}^{J+T}\left(\sum_{s\in A_r}\tau_{r-s}e^{\frac{V_s}{\mu_r}}\right)^{\mu_r}}\right]$ $0<\mu_m\leq 1,\ \sum_{m=1}^{J+T}\tau_m=1$	• Each alternative shares nest with T adjacent time periods • $\tau_0<\tau_1$ implies flight in time period 2 competes more with flight in time period 1 than 3
Cross-Nested Logit (CNL) Vovsha 1997 	$\sum_{m}\left[\frac{\left(\tau_{im}e^{V_i}\right)^{\frac{1}{\mu}}}{\sum_{j\in A_m}\left(\tau_{jm}e^{V_j}\right)^{\frac{1}{\mu}}}\times\frac{\left(\sum_{j\in A_m}\left(\tau_{jm}e^{V_j}\right)^{\frac{1}{\mu}}\right)^{\mu}}{\sum_{l}\left(\sum_{j\in A_l}\left(\tau_{jl}e^{V_j}\right)^{\frac{1}{\mu}}\right)^{\mu}}\right],$ $0<\mu\leq 1,\ \sum_{m}\tau_{jm}=1$	• Alternatives may appear in one or more nests • Logsums constrained to be equal
Generalized Multinomial Logit (Gen-MNL) Swait 2000 	$\sum_{m}\left[\frac{\left(\tau e^{V_i}\right)^{\frac{1}{\mu_m}}}{\sum_{j\in A_m}\left(\tau e^{V_j}\right)^{\frac{1}{\mu_m}}}\times\frac{\left(\sum_{j\in A_m}\left(\tau e^{V_j}\right)^{\frac{1}{\mu_m}}\right)^{\mu_m}}{\sum_{l}\left(\sum_{j\in A_l}\left(\tau e^{V_j}\right)^{\frac{1}{\mu_l}}\right)^{\mu_l}}\right],$ $0<\mu_m\leq 1,\quad 0<\tau<1$	• Alternatives may appear in two or more nests and logsums are estimated; however, each alternative must appear in exactly L nests due to constraint on allocation parameter

Table 4.3 ***Continued***

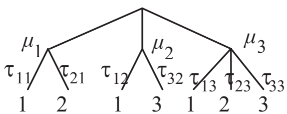
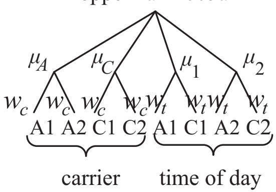
Generalized Nested Logit (GNL) Wen & Koppelman 2001  	$$\sum_m \left[\frac{\left(\tau_{im} e^{V_i}\right)^{\frac{1}{\mu_m}}}{\sum_{j\in A_m} \left(\tau_{jm} e^{V_j}\right)^{\frac{1}{\mu_m}}} \times \frac{\left(\sum_{j\in A_m} \left(\tau_{jm} e^{V_j}\right)^{\frac{1}{\mu_m}}\right)^{\mu_m}}{\sum_l \left(\sum_{j\in A_l} \left(\tau_{jl} e^{V_j}\right)^{\frac{1}{\mu_l}}\right)^{\mu_l}} \right]$$ $0 < \mu_m \leq 1, \sum_m \tau_{jm} = 1$	• Most general two-level GEV structure • All models shown above are special cases of GNL • Allocation and logsum parameters estimated • GNL is a special case of the Network GEV
Product Differentiation (PD) and Weighted Nested Logit (WNL) Breshahan *et al.* 1997, Coldren & Koppelman 2005a  	$$\sum_{f\in F} \left[\frac{e^{\frac{V_i}{\mu_f}}}{\sum_{k\in A_f} e^{\frac{V_k}{\mu_f}}} \times \frac{w_f \left(\sum_{k\in A_f} e^{\frac{V_k}{\mu_f}}\right)^{\mu_f}}{\sum_{c\in F} w_c \left(\sum_{s\in A_c} e^{\frac{V_s}{\mu_c}}\right)^{\mu_c}} \right]$$ $0 < \mu_f \leq 1, \sum_{c\in F} w_c = 1$	• Special case of GNL • Alternatives grouped by (allocated to) one or more observable characteristics • Exhibits unique structure in that each alternative appears in F nests (one for each dimension)

Table 4.3 ***Concluded***

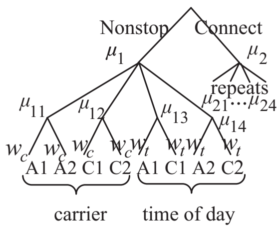
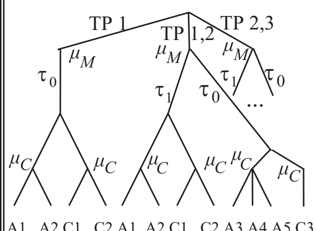
Nested-Weighted Nested Logit (N-WNL) Coldren & Koppelman 2005a 	$\frac{e^{\mu_m \Gamma_m}}{\sum_{l=1}^{M} e^{\mu_l \Gamma_l}} \times \sum_{f\in F}\left[\frac{e^{\frac{V_i}{\mu_{fm}}}}{\sum_{k\in A_f} e^{\frac{V_k}{\mu_{fm}}}} \times \frac{w_f e^{\frac{\mu_{fm}}{\mu_m}\Gamma_{fm}}}{\sum_{c\in F} w_c e^{\frac{\mu_{cm}}{\mu_m}\Gamma_{cm}}}\right], 0<\mu_{fm}\le\mu_m\le 1,$ $\sum_{f\in F} w_f = 1, \quad \Gamma_{fm} = \ln\left(\sum_{k\in A_f} e^{\frac{V_k}{\mu_{fm}}}\right), \quad \Gamma_m = \ln\left(\sum_{c\in F} w_c e^{\frac{\mu_{cm}}{\mu_m}\Gamma_{cm}}\right)$	• Extension of WNL (has one additional level); a NetGEV model • Alternatives first nested by one dimension (e.g. level of service), then by two or more dimensions • Alternatives with different higher-level dimension exhibit IIA
Ordered Generalized Extreme Value-Nested Logit (OGEV-NL) Coldren & Koppelman 2005a 	$\sum_{m=1}^{i+T}\left[\frac{e^{\mu_M\Gamma_m}}{\sum_{r=1}^{J+T} e^{\mu_M \Gamma_r}} \times \frac{e^{\frac{\mu_C}{\mu_M}\Gamma_{mc}}}{\sum_{d\in A_m} e^{\frac{\mu_C}{\mu_M}\Gamma_{md}}} \times \frac{\tau_{m-i} e^{\frac{V_i}{\mu_C}}}{\sum_{j\in A_{m,c}} \tau_{m-j} e^{\frac{V_j}{\mu_C}}}\right], 0<\mu_C\le\mu_M\le 1,$ $\sum_{m=1}^{T}\tau_m = 1, \quad \Gamma_{mc} = \ln\left(\sum_{j\in A_{m,c}} \tau_{m-j} e^{\frac{V_j}{\mu_C}}\right), \quad \Gamma_m = \ln\left(\sum_{c\in A_m} e^{\frac{\mu_C}{\mu_M}\Gamma_{mc}}\right)$	• Alternatives first grouped by time period in OGEV upper-level, then by one product dimension (e.g. carrier) in lower level. • Alternatives more than *T* time periods apart exhibit IIA

Table 4.4 Summary of direct- and cross-elasticities for select GEV models

Model	Direct Elasticity, $\eta_{X_{ik}}^{P_i}$	Cross Elasticity, $\eta_{X_{ik}}^{P_j}$
Multinomial Logit (MNL)	$(1-P_i)\,\beta_k X_{ik}$	$-P_i\beta_k X_{ik}$
Nested Logit (NL)	$\left[(1-P_i)+\left(\frac{1-\mu_m}{\mu_m}\right)(1-P_{i\|m})\right]\beta_k X_{ik}$	$-\left[P_i+\left(\frac{1-\mu_m}{\mu_m}\right)P_{i\|m}\right]\beta_k X_{ik}$
Paired Combinatorial Logit (PCL)	$\left[(1-P_i)+\sum_{j\neq i}\left(\frac{1-\mu_{ij}}{\mu_{ij}}\right)\frac{P_{i\|ij}P_{ij}(1-P_{i\|ij})}{P_i}\right]\beta_k X_{ik}$	$-\left[P_i+\left(\frac{1-\mu_{ij}}{\mu_{ij}}\right)\frac{P_{i\|ij}P_{j\|ij}P_{ij}}{P_j}\right]\beta_k X_{ik}$
Ordered Generalized Extreme Value (OGEV)	$\left[(1-P_i)+\sum_{m=i}^{i+T}\left(\frac{1-\mu_m}{\mu_m}\right)\frac{P_{i\|m}P_m(1-P_{i\|m})}{P_i}\right]\beta_k X_{ik}$	$-\left[P_i+\sum_{m=i}^{i+T}\left(\frac{1-\mu_m}{\mu_m}\right)\frac{P_{i\|m}P_{j\|m}P_m}{P_j}\right]\beta_k X_{ik}$
Cross-nested logit (CNL)	$\left[(1-P_i)+\sum_m\left(\frac{1-\mu}{\mu}\right)\frac{P_{i\|m}P_m(1-P_{i\|m})}{P_i}\right]\beta_k X_{ik}$	$-\left[P_i+\sum_m\left(\frac{1-\mu}{\mu}\right)\frac{P_{i\|m}P_{j\|m}P_m}{P_j}\right]\beta_k X_{ik}$
Generalized Multinomial Logit (Gen-MNL)	$\left[(1-P_i)+\sum_m\left(\frac{1-\mu_m}{\mu_m}\right)\frac{P_{i\|m}P_m(1-P_{i\|m})}{P_i}\right]\beta_k X_{ik}$	$-\left[P_i+\sum_m\left(\frac{1-\mu_m}{\mu_m}\right)\frac{P_{i\|m}P_{j\|m}P_m}{P_j}\right]\beta_k X_{ik}$
Generalized Nested Logit (GNL)	$\left[(1-P_i)+\sum_m\left(\frac{1-\mu_m}{\mu_m}\right)\frac{P_{i\|m}P_m(1-P_{i\|m})}{P_i}\right]\beta_k X_{ik}$	$-\left[P_i+\sum_m\left(\frac{1-\mu_m}{\mu_m}\right)\frac{P_{i\|m}P_{j\|m}P_m}{P_j}\right]\beta_k X_{ik}$
Product Differentiation (PD); Weighted Nested Logit (WNL)	$\left[(1-P_i)+\sum_{f\in F}\left(\frac{1-\mu_f}{\mu_f}\right)\frac{P_{i\|f}(1-P_{i\|f})P_f}{P_i}\right]\beta_k X_{ik}$	$-\left[P_i+\sum_{f\in F}\left(\frac{1-\mu_f}{\mu_f}\right)\frac{P_{i\|f}P_{j\|f}P_f}{P_j}\right]\beta_k X_{ik}$

Table 4.4 ***Concluded***

Nested-Weighted Nested Logit (N-WNL)	$\left[\begin{array}{l}(1-P_i) \\ +P_m \times \sum_{f\in F}\left[\left(\frac{1}{\mu_{fm}}-\frac{1}{\mu_m}\right)\frac{P_{i\|f,m}(1-P_{i\|f,m})P_{f\|m}}{P_i}\right] \\ -\left(\frac{1-\mu_m}{\mu_m}\right)\sum_{f\in F}P_{i\|f,m}P_{f,m}\end{array}\right]\beta_k X_{ik}$	$\left[\begin{array}{l}-P_i \\ -P_m \times \sum_{f\in F}\left[\left(\frac{1}{\mu_{fm}}-\frac{1}{\mu_m}\right)\frac{P_{i\|f,m}P_{j\|f,m}P_{f\|m}}{P_j}\right] \\ -\frac{1}{\mu_m}\frac{P_i}{P_j}\sum_{f\in F}P_{i\|f,m}P_{f,m} \\ +\frac{P_i}{P_j}\sum_{f\in F}P_{j\|f,m}P_{f,m}\end{array}\right]\beta_k X_{ik}$
Ordered Generalized Extreme Value-Nested Logit (OGEV-NL)	$\left[\left(\frac{1}{\mu_C}-P_i\right)+\frac{1}{P_i}\sum_{m=i}^{i+T}\left(\begin{array}{l}\left(1-\frac{1}{\mu_M}\right)P_m P_{c\|m}^2 P_{i\|c,m}^2 + \\ \left(\frac{1}{\mu_M}-\frac{1}{\mu_C}\right)P_m P_{c\|m} P_{i\|m,c}^2\end{array}\right)\right]\beta_k X_{ik}$	$\left[-P_i+\frac{1}{P_j}\sum_{m=j}^{j+T}\left(\begin{array}{l}\left(1-\frac{1}{\mu_M}\right)P_m P_{c\|m}^2 P_{i\|m,c}P_{j\|m,c} + \\ \left(\frac{1}{\mu_M}+\frac{1}{\mu_C}\right)P_m P_{c\|m} P_{i\|m,c}P_{j\|c,m}\end{array}\right)\right]\beta_k X_{ik}$

This page has been left blank intentionally

Chapter 5
Network GEV Models

Jeffrey P. Newman

Introduction

Chapter 4 provided an overview of the historical evolution of discrete choice models. Many of the models developed since the 1980's that have been applied to airline applications fall within the class of Generalized Extreme Value (GEV) models. GEV models are consistent with random utility theory, a methodology formalized by Manski (1977). GEV models relax the assumption that error terms are independently distributed while simultaneously maintaining the assumption that the total errors for each alternative are identically distributed. Conceptually, it is useful to further classify GEV models according to whether they have two levels, or more than two levels. The generalized nested logit (GNL) model, proposed by Wen and Koppelman in 2001, provides a general framework for analyzing GEV models that have two levels. The Network GEV (NetGEV) model, proposed by Daly and Bierlaire in 2006, provides a general framework for analyzing GEV models that have three or more levels. Although the GNL can be viewed as a special case of the NetGEV model, it is useful to distinguish between them, as the latter is relatively new to the literature and researchers are still investigating its theoretical properties.

This chapter provides an overview of the NetGEV model and highlights several of its key properties. Material from this chapter draws heavily on prior work by Newman (2008a, 2008b). Several techniques that researchers often use when developing new discrete choice models are highlighted. These techniques include the use of moment generating functions and the need to develop normalization rules to ensure the proposed model is uniquely identified. In the context of NetGEV models, two sets of normalization rules are introduced for models that satisfy a "crash free" or "crash safe" network property. The chapter concludes with a detailed example of a NetGEV model of airline itinerary choice based on synthetic data. The chapter concludes with an Appendix that provides an example that is used to illustrate how complex normalization rules can become if a general network structure is used. That is, although the NetGEV is a very flexible GEV model that can accommodate more general variance-covariance structures, it is important to recognize that the primary objective of NetGEV models—like all other discrete choice models—is to capture realistic competition structures across alternatives. Further, the use of intuitive competition structures tends to translate into networks with a well-defined pattern and straight-forward normalization rules.

For example, all of the two-level and three-level GEV itinerary choice models discussed in Chapter 4 satisfy both the crash free and crash safe network properties that will be introduced in this chapter.

Generating Functions for Generalized Extreme Value Models

It is not a trivial task to find joint extreme value distributions that relax the independence condition, while preserving the closed-form of model calculation. Fortunately, there is a family of such distributions, introduced by McFadden (1978), that can provide closed-form models. McFadden named this group the *Generalized Extreme Value* (GEV) family of distributions.

These distributions are created using *generating functions*. Such functions have to conform to much simpler criteria than the probability density functions (pdf) of the ultimate distributions. When a multivariate extreme value distribution is created using a valid generating function, it is certain to have a closed-form solution to calculate the resulting probabilities. The rules for a valid generating function $G(y)$ are:

- $G(y) \geq 0, \forall y \in \mathbb{R}_{+}^{J}$
- G is a homogeneous[1] of degree $\upsilon > 0$
- $\lim_{y_i \to +\infty} G(y) = +\infty, \quad i = 1,\dots,J$
- the mixed partial derivatives of G with respect to elements of y exist, are continuous, and alternate in sign, with non-negative odd-order derivatives, and non-positive even order derivatives.

To move from the generating function to a discrete choice model, y in the generating function is replaced with $\exp(V)$. The resulting choice model has a closed-form probability expression, and is consistent with random utility maximization theory. Formally, the probability associated with alternative i is derived from the generating function as follows:

$$P_i = \frac{y_i G_i(y)}{G(y)}$$

where $G_i(y)$ is the first derivative of G with respect to y_i. Different generating functions will result in different probability density functions within the Generalized Extreme Value family. The primary benefit of varying the generating function is that different generating functions will result in multivariate density

1 McFadden (1978) originally required that G had to be homogeneous of degree 1, but this condition was relaxed by Ben-Akiva and Francois (1983), such that G needs only be homogeneous of any positive degree.

functions with different attributes, in particular with different covariance matrices. The ability to incorporate covariance between the random portion of utility allows the modeler to partially account for relationships between alternatives that are not expressed in the observed characteristics of those alternatives.

Since the development of the GEV structure for discrete choice models in 1978, substantial efforts have been put forth to find new forms of GEV model, exhibiting more varied covariance structures. Progress was initially slow. Although the criteria for a generating function are simpler than those for a multivariate extreme value function, the fourth point (alternating signs of partial derivatives) is still generally not easy to check for most functional forms. For some time, modelers were limited to the initial multinomial logit (MNL) and nested logit (NL) models, which both predated the more general GEV formulation. Ultimately, Wen and Koppelman (2001) proposed the generalized nested logit (GNL) model, a more general form that encompasses all previous such models, with the exception of the multi-level NL model. Using the notation introduced in earlier chapters and suppressing the index of n for individual for notational convenience, the generating functions for the MNL, two-level NL, and GNL functions are given, respectively, as:

MNL: $G(y) = \sum_{j \in C} y_j$

Two-level NL: $G(y) = \sum_{m=1}^{M} \left(\sum_{i \in A_m} y_i^{1/\mu_m} \right)^{\mu_m} \quad 0 < \mu_m \le 1,\ i \in A_m,\ m = 1,\ldots,M$

GNL: $G(y) = \sum_{m=1}^{M} \left(\sum_{i \in A_m} \left(\tau_{im} y_i \right)^{1/\mu_m} \right)^{\mu_m} \quad \tau_{im} \ge 0,\ \sum_{m=1}^{M} \tau_{im} = 1 \forall i$

The GNL, unlike the NL model, is limited to only a single level of nests, and does not allow hierarchical[2] (or multi-level) nesting.

Beyond the need to ensure that the mathematical forms of generating functions were compliant with the GEV rules, the process of discovering new GEV models was hampered by the availability of computing power. More complex GEV forms, such as the GNL model, require more computations to calculate the resulting model probabilities and parameters, especially in light of the fact that there are generally more parameters in such models. Even though this computational effort is low compared to numerical integration, it can still be large compared to MNL and NL models. Technological advancements in computing power and data storage have thus made it possible to estimate ever more detailed and complex models. For example, Coldren and Koppelman (2005a) introduced a three-level weighted nested logit model (WNL) as well as a nested-weighted nested logit model

2 Note that "hierarchical" refers to a multi-level nesting structure and associated variance-covariance structure. It *does not* refer to a sequential decision process.

N-WNL). These models are specific instances of the more general NetGEV model, proposed by Daly and Bierlaire (2006).

Network GEV

The NetGEV uses a topological network of links and nodes to stitch together sub-models into one complete discrete choice model. Each sub-model represents a GEV model that includes only a portion of the choice set. By progressively connecting these sub-models, the whole choice set is eventually represented in one final model, which is by construction still a correct GEV form.

To create a NetGEV model, one begins with a network. It is very similar to the graphical representations of the models that have been discussed in previous chapters. Formally, the network must be finite, directed (each link connects *from* one node *to* another), connected (between any pair of nodes, there is a path between them along links, regardless of the links' direction), and circuit-free (there is no a directed path along links from any node back to itself). The network has one source or root node, which only has outgoing links, as well as a sink node (with only incoming links) to represent each discrete alternative.

Using a slight simplification of Daly and Bierlaire's (2006) network, as detailed in Newman (2008b), first start at the bottom of the network, with the elemental alternatives. At each alternative node i, a sub-model is created where the generating function $G(y) = y_i = \exp(V_i)$. The model is very simple, and it trivially conforms to all the necessary conditions, given that it applies to a subset of alternatives that contains only one alternative (i).

For the other nodes in the network (including the root node), the model for each node is assembled from the models at the end of each of the outbound links, according to the formula:

$$G^i(y) = \left(\sum_{j \in i^{\downarrow}} \left[\left(a_{ij} G^j(y) \right)^{1/\mu_i} \right] \right)^{\mu_i} \tag{5.1}$$

where:

i is the relevant node,

$i^{\downarrow}$ is the set of successor nodes to i (the nodes at the end of outbound links),

a_{ij} is an allocation parameter associated with each link in the network,

μ_i is a scaling parameter associated with node i.

Note that at this point in the discussion, new notation has been introduced to facilitate the discussion of the NetGEV model. Specifically, although the scaling parameters (and their associated normalization rules) are similar to the interpretation of logsum coefficients seen earlier in the context of NL models, a new notation for allocation parameters associated with the links of the network

has been introduced (namely a_{ij}). These a allocation parameters are distinct from the τ allocation parameters discussed in Chapter 4 in the sense that the a allocation parameters are now more general. That is, a and τ are functionally equivalent; however, in the general NetGEV framework, it could be necessary to impose some complicated non-linear constraints on the values of τ. To simplify these constraints, the parameter will be transformed, and the notation a will be used in place of τ, where a is a function of α_{ij}, to underscore the distinction. In addition, instead of associating nodes with a specific level in a tree, a set of nodes for the entire NetGEV structure, N, has been defined. Some of the nodes represent elemental alternatives, whereas other nodes represent intermediate nodes or the root node. Thus, the NetGEV model, in addition to the $\boldsymbol{\beta}$ parameters embedded inside the systematic utility (V_i) of each node, also has an a parameter on each network link, and a μ parameter on each network node, excluding the elemental alternative nodes.

There are a few constraints of the value of the parameters. Each a parameter must be greater than zero. If any a is equal to zero, that is the equivalent of deleting the associated link from the network, which is acceptable as long as the network remains connected. Each μ parameter must be positive, and smaller than the μ parameters of all predecessor nodes (those that are at the other end of incoming links). Additionally, in order to be identified in a model, these parameters need to be normalized, similar to $\boldsymbol{\beta}$ parameters in a utility function (where one alternative specific constant is normalized to be equal to zero). This can be done by setting one μ and one a to a specific value. For μ parameters, usually the root node μ is set equal to 1. For a parameters, the normalization can be done in various different ways, and the ideal method will vary with the structure of the network.

The relationship between G^i and V_i, the systematic utility of the alternative, is simple when node i is an elemental alternative, i.e., $G^i = \exp(V_i)$. It is useful to conceptualize a similar relationship between G^n and V_n for nesting nodes, even though those nodes do not have a direct systematic utility *per se*. V_n for nesting nodes is the logsum of the nest, which is a relevant measure of utility. In the NL model, V_n is the scale adjusted logsum value for the nest. It retains a similar function in the NetGEV structure.

Advantage of NetGEV

The NetGEV model is more flexible than other GEV models, including the GNL model, as it is able to represent a greater range of possible correlation structures between alternatives. In particular, the hierarchical nesting structure allows strongly correlated alternatives to still be loosely correlated with other alternatives. Wen and Koppelman (2001) begin to explore the differences between the GNL and the hierarchical form as expressed in the NL model. They conclude that the GNL can generally approximate an NL model. The NetGEV model, on the other hand, can close that gap entirely.

For example, consider the famous red bus/blue bus problem. In the traditional scenario, a decision-maker is initially faced with a choice between travelling in a car or in a red bus, as in the *A* model in Figure 5.1. In the simplest case, these alternatives are considered equally appealing, and each has a 50 percent probability of being chosen. When a new blue bus alternative is introduced, which is identical in every way to the red bus, one would expect the bus riders to split across the buses, but car drivers would not move over to a bus alternative. In the MNL model, however, this does not happen. Instead, as in the *B* model in Figure 5.1, the buses draw extra probability compared to the original case. The introduction of the NL model, as in the *C* model, allows the error terms for the bus alternatives to be perfectly correlated, and the expected result is achieved.

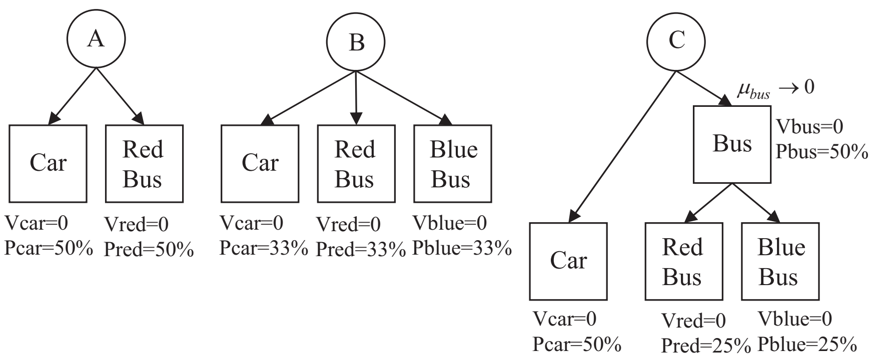


Figure 5.1 One bus, two bus, red bus, blue bus

Source: Adapted from Newman 2008a: Figure 2.1 (reproduced with permission of author).

However, in a revised scenario, the original case is not binary choice, but instead it is a three-way choice, between a car, a bus, and a train. Further, the initial model can be constructed as a GNL model (shown in the *D* model in Figure 5.2), so that the car and bus alternatives are partially nested together (both get stuck in traffic), and the bus and train alternatives are also partially nested together (both are mass transit). In this model, the utility of the bus tends to fall between car and train, so that its probability is slightly reduced relative to the others. Again, the blue bus is introduced into the market, identical to the red bus. If the blue bus is inserted into the GNL model with the same nesting setup as the existing red bus, as in the *E* model in Figure 5.2, the probabilities of the car and train alternatives are adversely affected. A new "bus" nest could be introduced to induce the required perfect correlation between the error terms of the buses, but under the constraints of the GNL model, the allocations of the buses to the traffic and transit nests would need to be reduced (to zero), eliminating the correlation between the buses and the other alternatives.

The NetGEV model removes that constraint of the GNL model, and allows hierarchical nesting, as in a standard NL model. Thus, the nesting structure in the

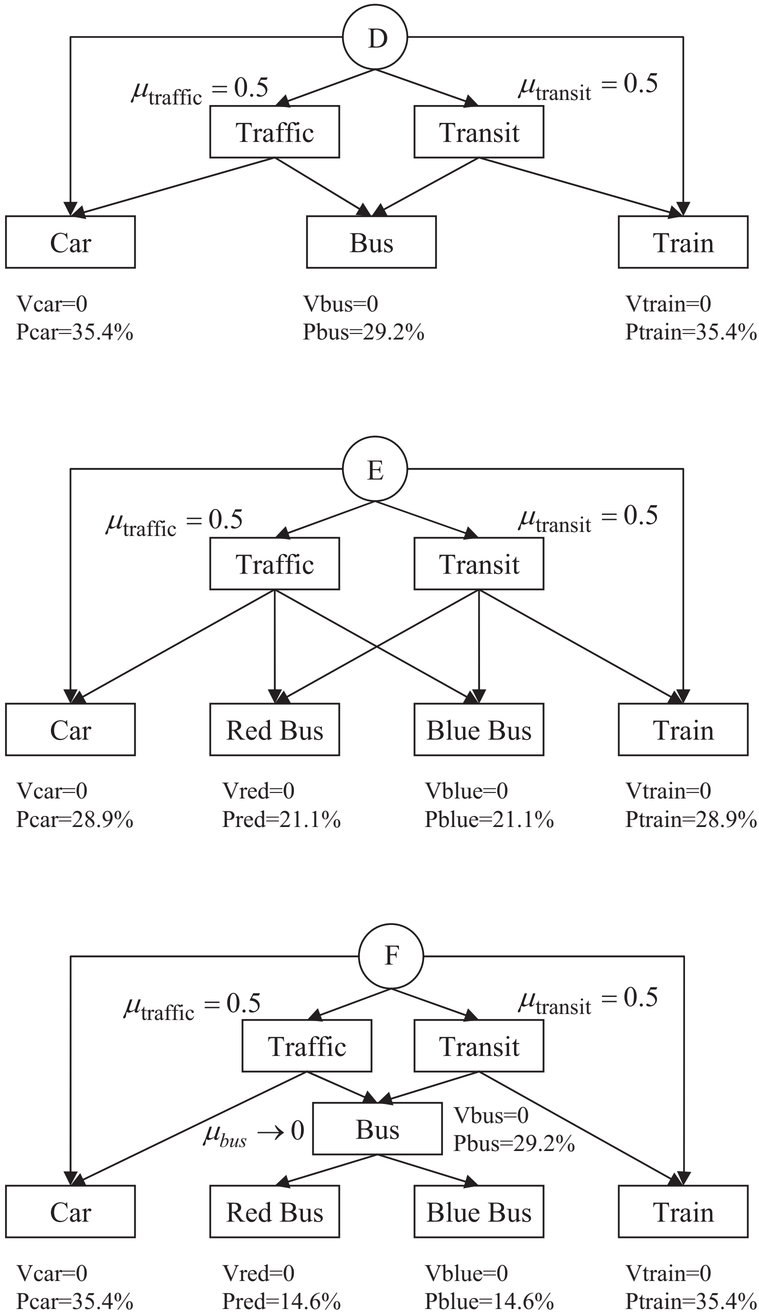


Figure 5.2 The blue bus strikes again

Source: Adapted from Newman 2008a: Figure 2.2 (reproduced with permission of author).

F model of Figure 5.2 can be created, linking together the buses before allocating them to traffic and transit nests. The probabilities for car and train can be preserved, with the red and blue buses splitting the bus market only.

Normalization of Parameters

The NetGEV model as formulated is over-specified, so that is not possible to identify a unique likelihood maximizing set of parameters. The over-specification is similar to that observed in attempts to maximize $f(x, y, z) = -(x + y)^2 + (z/z)$. This problem cannot be solved to an identifiable unique solution; any value for any individual parameter can be incorporated into a maximizing solution. Some parameters are unidentified as a set (as are x and y), and can only be identified if one of the set is fixed at some externally determined value (e.g., setting $y = 1$) or if some externally determined relationship is applied (e.g., setting $x = y$). Other parameters are intrinsically unidentified (in this example, z), and cannot be identified at all.

Mathematically, this is expressed in the derivatives of f with respect to its parameters. The first derivative of f with respect to an intrinsically unidentified parameter is globally zero. Parameters unidentified in sets can individually have calculable first partial derivatives, but the Hessian matrix of second derivatives is singular along the ridge of solutions.

In the NetGEV model, over-specification (and the resulting need for normalization conditions) can arise for multiple reasons. Earlier, the need to normalize logsum and allocation parameters in the context of NL and GNL models was discussed. The NetGEV model also needs normalization rules for logsum parameters (which are similar to the rules developed in the context of NL models) and allocation parameters (which are now dependent on the underlying network structure). Additional normalization constraints are also needed to handle over-specification caused by the topological structure of the GEV network.

Topological Reductions

The topographical structure of the GEV network can create over-specification, by including extraneous nodes and edges that do not add useful information or interactions to the choice model. Fortunately, these extraneous pieces can be removed from the network without changing the underlying choice model. Figure 5.3 provides a pictorial representation of the extraneous nodes and edges covered in this subsection.

Degenerate nodes A degenerate node is a node in the network that has exactly one successor. The G function for a degenerate node d collapses to a single term:

$$G^d(y) = \left(\sum_{j \in d^{\downarrow}} \left[\left(a_{dj} G^j(y) \right)^{1/\mu_d} \right] \right) = a_{dj} G^j(y) \tag{5.1}$$

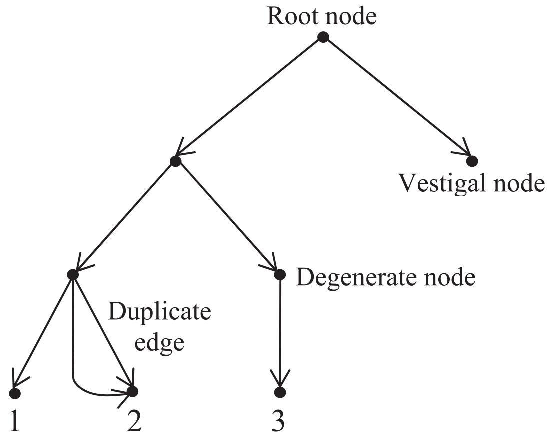


Figure 5.3 Network definitions

In this case, μ_d drops out of the equation, and has no effect on G^d, and thus no effect on any other G in the network, including G^R. Since μ_d disappears from the calculation, it is intrinsically unidentified. Degenerate nodes can be removed from the network, or their associated parameters can be fixed at some value. Although certain non-normalized NL models may require degenerate nodes to correctly normalize the model (see Koppelman and Wen 1998a), the NetGEV model does not require such nodes.

Vestigial nodes A vestigial node is a node which has no successors, but is not associated with an elemental alternative. Although such nodes generally would not be expected in any practical application, the definition of a GEV network does not technically preclude their existence. The G function for such a node would always equal zero, as the set of successor nodes in the summation term of Equation 5.1 is empty. The removal of such nodes from the network would obviously not affect the resulting choice probabilities. As with degenerate nodes, if they are not removed, it will be necessary to externally identify the value of their logsum parameters.

Duplicate edges Duplicate edges also add complexity to the network without providing any useful properties. A duplicate edge is any edge in the network that shares the same pair of ends as another edge. As the network is defined to be circuit free, all duplicate edges will always be oriented in the same direction. The allocation parameters on any set of duplicate edges are jointly unidentified, but the extra edges can be removed without altering the underlying choice model.

When a GEV network has been stripped of degenerate and vestigial nodes, and duplicate edges, it can be considered a *concise GEV network.* Each of these processes results in the removal of nodes or edges from the network, and since any GEV network is finite, the process of reducing any GEV network to its equivalent

concise network must conclude after a finite number of transformations. As it is not restrictive to do so, the remainder of this chapter will assume that GEV networks are concise.

Normalization of Logsum Parameters

It is well known that it is necessary to normalize logsum parameters in NL models, as the complete set of logsum parameters is over-specified (Ben-Akiva and Lerman 1985). As the NetGEV model is a generalization of the nested logit model, it follows that the logsum parameters in this model will also need to be normalized. In particular, as mentioned by Daly and Bierlaire (2006), the logsum parameters are only relevant in terms of their ratios. This is not quite as obvious in the mathematical formulation presented here as it is in the original formulation, but since they are equivalent the condition still holds. Setting the logsum parameter for any single nest (except the nodes associated with elemental alternatives, and degenerate nests) to any positive value will suffice to allow the remaining logsum parameters to be estimated. Typically, it will be convenient to fix the logsum parameter of the root node equal to one.

The logsum parameters of degenerate nodes (and elemental alternatives) are intrinsically unidentifiable, and thus cannot be used as anchors to identify the parameters on other nodes. If any degenerate node is not removed from the network, then the associated logsum parameter must be set externally.

Normalization of Allocation Parameters

It is also necessary to normalize the allocation parameters in a NetGEV model. Multiplying all the a values in Equation 5.1 by a constant is equivalent to multiplying the G function by the constant, which does not change a GEV model. More generally, for any network cut that divides the root node from all alternative nodes, multiplying all the a values for all edges in the cut by a constant is equivalent to multiplying G^R by that constant. This change would not affect the ratio of G^R and its derivatives with respect to y, and thus would not affect the resulting probabilities of the model. In order to be able to estimate the allocation parameters, some relationships between them must be fixed externally.

The imposition of these relationships between allocation parameters could potentially create an undesired bias in the model. An unbiased model is one such that the expected value of the random utility for any alternative i is equal to the systematic (observed) utility for that alternative, plus a constant with fixed value regardless of the alternative:

$$\bar{U}_i = V_i + \bar{\varepsilon}_i = V_i + \xi \tag{5.2}$$

and thus $\bar{\varepsilon}_i = \xi$. An unbiased model does not imply that actual observed choice preferences will not be biased in favor of one or more alternatives, but rather

indicates merely that a model will not over- or under-predict the probability of an alternative due only to the structure of the model.

The constant expected value of ε_i in Equation 5.2 only applies to elemental alternatives. Although the log of the generating function G may create a value V that is analogous to the systematic utility of an elemental alternative, there is no explicit error term ε for a nesting node. If one were to be assumed, its expected value could be any value, not necessarily ξ.

To ensure the unbiased condition is met, the normalization of a will depend on the topographical structure of the network. Normalizations for two topographical structures are presented in this chapter: one for networks that are crash free and one for networks that are crash safe. The appendix to this chapter contains an example of one method a researcher can use to normalize a network that is neither crash free nor crash safe. The example is used to highlight how normalization rules for the allocation parameters can become much more complex, even when seemingly minor changes are made to a network structure.

Before presenting the normalization rules for allocation parameters for crash free and crash safe networks, it is helpful to visualize what is driving the need

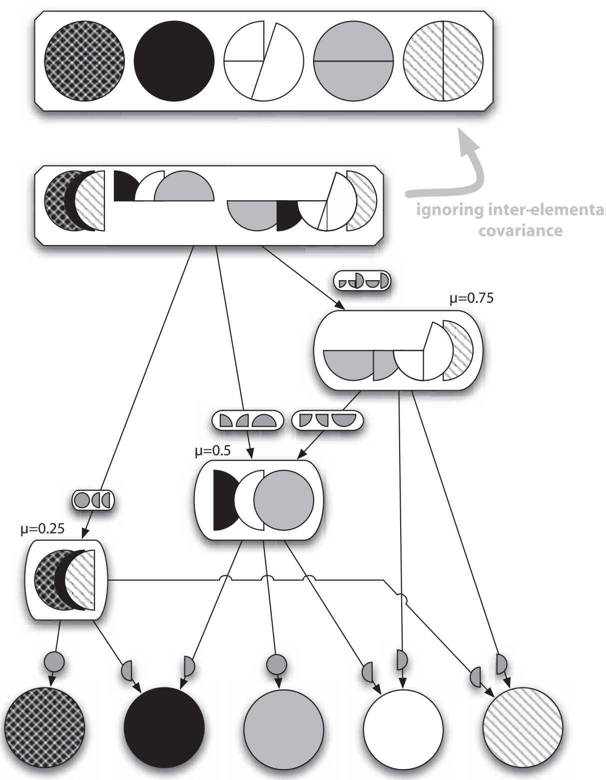


Figure 5.4 Ignoring inter-elemental covariance can lead to crashes

to normalize these parameters in the first place. Figure 5.4 presents a network with five elemental alternatives, (represented at the lowest level of the tree by nodes with different shading and backgrounds). Consider the second elemental alternative from the left, and assume that ½ is allocated to the 0.25 node and ½ is allocated to the 0.5 node. This is represented by the dark left half-circle and the dark right half-circle at the 0.25 and 0.50 nodes, respectively. Moving further up the tree, the 0.25 node connects directly to the root, so the entire ½ circle is allocated to the root node. However, there are two paths to reach the root from the 0.5 node—one that is direct and one that goes through the 0.75 node first. Assuming ½ of the alternative is allocated to each path, a ¼ circle arrives to the root directly from the 0.5 node and a ¼ circle arrives to the root through the path that goes through the 0.75 node. At the root, all of the pieces recombine and sum to one. That is, loosely speaking, the variance components associated with the second (darkest) alternatives remain intact as they travel up the network, and sum to one at the root node.

The core problem occurs with a situation depicted with the fourth (lightest) alternative from the left. In this case, ½ of the alternative is allocated to the 0.5 node and ½ is allocated to the 0.75 node. From the 0.5 node, ¼ of the circle goes to the root and ¼ goes to the 0.75 node. The problem occurs at the 0.75 node, in that pieces of the same alternative are being recombined prior to reaching the root. In this case, the total variance components or circle associated with the 0.75 node is less than its allocations, i.e., less than ¾ of a circle, as the two paths are "perfectly correlated" for this alternative. Stated another way, a "crash" has occurred at an intermediate node as pieces of the same alternative arrive from different paths. In this case, normalization rules (which can be loosely thought of as "airbags") are needed to ensure all of the pieces are properly recombined and full circles are represented at the root node.

Conceptually, this example serves to highlight another problem that can occur when creating general network models—they may not be fixable. That is, the network structure itself may lead to over-identification and the only way to successfully estimate parameters is to change the underlying network structure. Although theoretically, this will lead to an altered variance-covariance matrix (and different model with potentially different choice probabilities), in practical terms, the author hypothesizes that it will be difficult to justify networks such as the one in Figure 5.4 from a behavioral perspective. That is, the majority of behavioral-realistic inter-alternative competition structures follow fairly straight-forward network structures. Two network structures that have been most frequently encountered in the aviation airline context (and include all of the itinerary choice models presented to date) include: 1) networks that exhibit the crash free property; and/or 2) networks that exhibit the crash safe property. Normalization rules for both of these network structures that have been published in the literature (Newman 2008b) are discussed below. It is important to note, however, that the rules provided here are only one of many possible set of rules. Investigation of the theoretical properties of the NetGEV model remains an active area of research.

Crash free networks A crash free network is one where multiple pieces of the same alternative are re-combined only at the root node. That is, for any node $i \in C$, no two distinct paths leading from R to i may share the edge connected to R. All paths must diverge separately from the root node, and although they may converge sooner than reaching the elemental alternative node, they may not share an edge that emanates from the root node and subsequently diverges.

For example, the network on the left side of Figure 5.5 does not conform to this criterion, because elemental alternative C has multiple path divergence points on paths from R. There are four distinct paths through the network from R to C: $R \rightarrow M \rightarrow C$, $R \rightarrow K \rightarrow C$, $R \rightarrow K \rightarrow N \rightarrow C$, and $R \rightarrow N \rightarrow C$. The paths $R \rightarrow K \rightarrow C$ and $R \rightarrow K \rightarrow N \rightarrow C$ share a common edge emanating from R, which is not allowed. The network on the right side of Figure 5.5 is similar to the network on the left, with the only difference being that the edge from K to C is missing, eliminating the path $R \rightarrow K \rightarrow C$. Of the three remaining paths, no two share an edge emanating from R. This reduced network is crash free. Note that the crash free network in Figure 5.5 is functionally different from the original network, and removing an edge from a network can potentially result in a radically different model. (A strategy to adjust a nonconforming network is examined in the Appendix of this chapter.)

In a crash free network, for any node except the root node there can be at most one unique path from that node to any other node. If there were more than one path from any node i other than the root node to any other node, then those multiple paths could be extended backwards from i to the root node, sharing common edges, including the edge connecting to the root. Checking this criterion requires building a directed tree from each node connected directly to the root node. If any node in the completed tree has any outbound edges that are not included in the tree, then

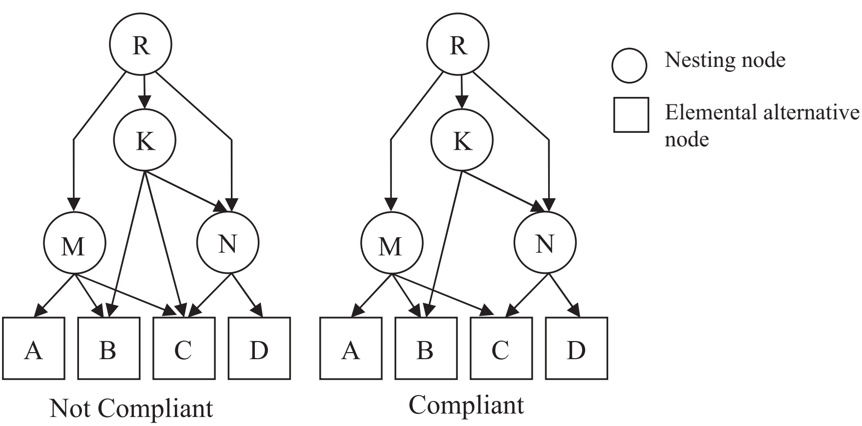


Figure 5.5 Making a GEV network crash free

Source: Adapted from Newman 2008b: Figure 2 (reproduced with permission of Elsevier).

that edge must connect to another node in the tree, completing a second path to that node, and violating the crash avoidance criterion. Multiple paths diverging from nodes not directly connected to the root node will be captured in the tree(s) of that node's predecessor(s) in the set of nodes connected to the root.

As shown in Newman (2008b), when a GEV network is crash free, setting the allocation terms $a_{ij} = \alpha_{ij}^{\mu_R}$ and enforcing $\sum_{i \in j^{\uparrow}} \alpha_{ij} = 1$ will ensure unbiased error terms. However, the crash avoidance restriction is not the only way to allow an unbiased normalization of the allocation parameters in a NetGEV model.

Crash safe networks Crash safe normalization imposes a slightly different restriction on the graph that defines the NetGEV model: for any node $i \in C$, no two distinct paths leading from R to i may share the edge connected to i. That is, all paths must converge separately at the elemental alternative node, and although they may diverge later than departing the root node, they may not share an edge arriving at the elemental alternative node.

This condition is easier to check than crash avoidance, as only elemental alternative nodes can have multiple predecessor nodes. Since the network is connected and has only one root node without predecessors, every node in the network must have at least one path connecting to it from the root node. If any node j has more than one predecessor node, then it must also have more than one possible path from the root node, as there must be at least one path through each of the predecessor nodes. Those paths would then converge at j. If j is not an elemental alternative node, then the condition for crash safety would be violated.

For example, the network on the left side of Figure 5.6 does not conform to this criterion, because elemental alternative C has multiple path convergence points. There are three distinct paths through the network from R to C: $R \rightarrow M \rightarrow C$, $R \rightarrow K \rightarrow M \rightarrow C$, and $R \rightarrow K \rightarrow N \rightarrow C$. The paths $R \rightarrow M \rightarrow C$ and $R \rightarrow K \rightarrow M \rightarrow C$ share a common edge terminating at C, which is not allowed. The network on the right side of Figure 5.6 is the same, except the edge from K to M is missing, eliminating the path $R \rightarrow K \rightarrow M \rightarrow C$. The two remaining paths do not share an edge terminating at C. This reduced network is crash safe. Again, the two networks shown in Figure 5.6 represent two different models, with potentially different probabilities for alternatives.

The normalization of a network with this topology is different from that described for crash free networks. Instead of ensuring that partial allocations of alternatives recombine at the root node (and thus without any internal correlation), the partial alternatives are allowed to recombine at any arbitrary location, with possibly some correlation between the partial alternative's error terms. However, the location of the distribution of the partial alternative error terms is augmented, so that the location of the recombined error distribution will still be constant across alternatives.

In order to provide a general algorithm to ensure this augmentation can be done correctly for each alternative without conflicting with the necessary corrections for other alternatives, all of the splitting of partial alternatives under this topological condition is done on the edges connecting to the elemental

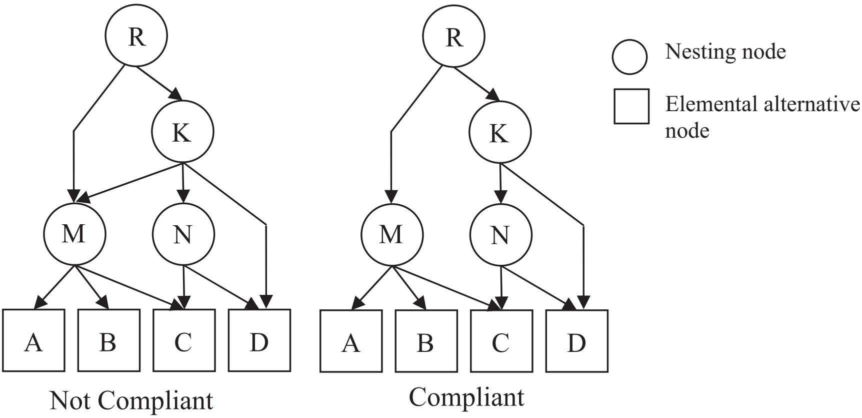


Figure 5.6 Making a GEV network crash safe

Source: Adapted from Newman 2008b: Figure 3 (reproduced with permission of Elsevier).

alternatives. Each allocation parameter on these edges is associated with one and only one elemental alternative, so that each alternative's partial alternatives can be adjusted independently. It is not necessary that a network is crash safe in this way in order to achieve an unbiased normalized model, if multiple alternatives are constrained such that the necessary adjustments on nesting nodes do not conflict, but it is sufficient and convenient if the criterion described here holds.

The crash safe normalization is more complex than the crash free method, and will require the introduction of some new network descriptors. As described earlier, each node in *N,* excluding *R*, has exactly one predecessor. For any node *n* in *N,* let $\overleftarrow{n}$ be the predecessor of *n*, $\overleftarrow{\overleftarrow{n}}$ the predecessor of $\overleftarrow{n}$, $\overleftarrow{\overleftarrow{\overleftarrow{n}}}$ the predecessor of $\overleftarrow{\overleftarrow{n}}$, and so on backwards through the network until $\mathring{n}$, which is eventual predecessor of *n* and an immediate successor of *R*. For each elemental alternative node *i*, let G^i be a sub-graph constructed of only the nodes and edges that have *i* as an eventual successor, excluding *i* itself.

If $a_{jk} = 1$ for all *k* in *N*, then the allocation parameter for the edge connecting from any node in *N* to a node *i* in *C* can also be considered as the allocation $\vec{\alpha}_{PR_i}$ to the entire path PR_i from *R* to *i* that uses that edge.

For each node *j* in *N* define $T(R, j, i)$ as the set of all paths from *R* to *i* that pass through *j*, and $\tilde{\alpha}_{R_{ji}}$ as the total allocation to those paths:

$$\tilde{\alpha}_{Rji} = \sum_{p \in T(R,j,i)} \vec{\alpha}_{pRi}$$

or alternatively:

$$\tilde{\alpha}_{Rji} = \alpha_{ji} + \sum_{k \in \left\{G^i \cap j^{\downarrow}\right\}} \vec{\alpha}_{Rki}$$

For a GEV network which is crash safe as described above, setting $\alpha_{jk} = 1$ for all k in N and

$$a_{ni} = \left(\alpha_{ni}\right)^{\mu_n} \left(\tilde{\alpha}_{Rni}\right)^{\mu_{\dot{n}}-\mu_n} \left(\tilde{\alpha}_{R\dot{n}i}\right)^{\mu_{\ddot{n}}-\mu_n} \left(\tilde{\alpha}_{R\ddot{n}i}\right)^{\mu_{\dddot{n}}-\mu_{\ddot{n}}} \ldots \left(\tilde{\alpha}_{R\tilde{n}i}\right)^{\mu_R-\mu_{\tilde{n}}}$$

for all i in C, or equivalently,

$$a_{ni} = \left(\frac{\alpha_{ni}}{\tilde{\alpha}_{Rni}}\right)^{\mu_n} \left(\frac{\tilde{\alpha}_{Rni}}{\tilde{\alpha}_{R\dot{n}i}}\right)^{\mu_{\dot{n}}} \left(\frac{\tilde{\alpha}_{R\dot{n}i}}{\tilde{\alpha}_{R\ddot{n}i}}\right)^{\mu_{\ddot{n}}} \left(\frac{\tilde{\alpha}_{R\ddot{n}i}}{\tilde{\alpha}_{R\dddot{n}i}}\right)^{\mu_{\dddot{n}}} \ldots \left(\frac{\tilde{\alpha}_{R\tilde{n}i}}{\tilde{\alpha}_{RRi}}\right)^{\mu_R}$$

and enforcing $\sum_{j\in i^{\uparrow}} \alpha_{ji} = 1$ will ensure unbiased error terms.

Bias constants If neither topological condition applies to a GEV network, it is still possible to normalize the allocation parameters and retain an "unbiased" model. One way to do this is to include a complete set of alternative specific constants (except for one arbitrarily fixed reference alternative) in the model. This method does not ensure unbiased systematic utility through constant expected value for the error terms as in Equation 5.2. Instead, $\bar{\varepsilon}_i$ is allowed to vary from κ, but the necessary adjustment $(\bar{\varepsilon}_i - \kappa)$ is incorporated into V_i itself. Unfortunately, this is undesirable because it conflates the model bias correction with the actual choice preference bias. This can cause problems in interpreting these model parameters, and in comparing the parameters between models, even when those models are estimated with the same underlying data. Additionally, there are various reasons why it might be undesirable to include a complete set of alternative specific constants in a model, often because the number of alternatives can be vast for complex models.

Disaggregation of Allocation

Relaxing Allocation Parameter Constraints

As discussed earlier, the normalization of the NetGEV model requires that the allocation parameters sum to a constant independent of the source node, typically 1. In either the crash safe or crash free conditions, the necessary constraint is $\sum_{j\in i^{\uparrow}} \alpha_{ji} = 1$. Imposing this restriction directly on estimated parameters results in additional complications, as the parameters are bounded not only by fixed values but also by each other. However, this restriction can be relaxed by transforming the parameters using the familiar logit structure:

$$\alpha_{ji} = \frac{\exp\left(\phi_{ji}\right)}{\sum_{k\in i^{\uparrow}} \left[\exp\left(\phi_{ki}\right)\right]} \tag{5.3}$$

Under this transformation, a new set of ϕ parameters replaces the α parameters throughout the network on a one-for-one basis. Instead of the requirement that the α parameters' add up to one among the set of parameters associated with each node with more than one predecessor, the ϕ parameters may vary unbounded across ¡ so long as one ϕ in each such group is fixed to some constant value (typically zero). This is a significant advantage in parameter estimation, as nonlinear optimization algorithms are substantially easier to implement when there are no (or fewer) constraints on the parameters.

Subparameterization of Allocation

Replacing the α parameters with a logit formulation not only simplifies the process of estimating the allocation parameters, it also opens up the possibility creating a much richer model. The logit structure for nest allocation allows for the incorporation of data into the correlation structure of error terms:

$$\alpha_{tji} = \frac{\exp\left(\phi^*_{ji} + \phi_{ji} Z_t\right)}{\sum_{k \in i^{\uparrow}} \left[\exp\left(\phi^*_{ki} + \phi_{ki} Z_t\right)\right]} \tag{5.4}$$

where ϕ^*_{ji} is the baseline parameter as in Equation 5.3, Z_t is a vector of data specific to decision-maker t, and ϕ_{ji} is a vector of parameters to the model which are specific to the link from predecessor node j to successor node i. Assuming that the first value in Z_t is 1 (defining a "link-specific" constant), Equation (5.4) can be simplified to:

$$\alpha_{tji} = \frac{\exp\left(\phi_{ji} Z_t\right)}{\sum_{k \in i^{\uparrow}} \left[\exp\left(\phi_{ki} Z_t\right)\right]} \tag{5.5}$$

Thus, the G function for nesting nodes becomes

$$G^i(y) = \left(\sum_{j \in i^{\downarrow}} \left(\frac{\exp\left(\phi_{ji} Z_t\right)}{\sum_{k \in i^{\uparrow}} \left[\exp\left(\phi_{ki} Z_t\right)\right]} G^j(y) \right)^{1/\mu_n} \right)^{\mu_n}$$

This then results in a heterogeneous covariance network GEV model (HeNGEV). The heterogeneity is created by the ϕ parameters, which relate the allocations of nodes to predecessor nests to the attributes of the decision-makers. Because the data elements in Z_t are all tied to the decision-makers (and cannot vary by node),

all of the ϕ parameters are all link specific parameters, analogous to alternative specific parameters in an MNL model. As usual for "alternative" specific constants and variables logit models, one of the vectors ϕ_{ji} must be constrained to some arbitrary value, usually zero. The remaining ϕ vectors can vary unconstrained in both positive and negative regions of ¡ . By changing the allocations of nodes in response to decision-maker attributes, the model can react not only in determining the systematic (observed) utility, but also in determining the correlation structure for random (unobserved) utility. This model thus allows both the amount and the form of covariance to vary across decision-makers.

For example, consider an air itinerary and fare class choice model, built on a network model. The network is bifurcated into two substructures, one with itinerary nested inside fare class, and the other with fare class nested inside itinerary. Each particular potential ticket choice is partly allocated to both substructures. The allocation parameters could then vary based on frequent flyer status, with program member decision-makers tending to choose based on one substructure, and nonmember decision-makers tending to choose based on the other.

Since the form of Equation 5.5 is by construction strictly positive, the HeNGEV model already meets one of the conditions of the NetGEV formulation, that a is positive. As long as the non-increasing μ parameters condition also holds, the HeNGEV model will be consistent with utility maximization.

Application

The HeNGEV model, by its nature, is most useful for analyzing complex decisions. Choices where decision-makers only have a small handful of options do not provide a lot of opportunity for complex correlation structures. In complex choices with large choice sets, the benefits of this flexible model can become more apparent. One typical such decision occurs in travel booking, where travelers must choose among a variety of itineraries when selecting an airline ticket. A hypothetical choice scenario is used to illustrate the model.

Data Generation

This scenario involves data that would approximate what might be observed for a flight itinerary choice between two medium sized airports in the United States. There are a variety of itinerary options (nonstop, single connection, and double connection flights on five different carriers) within a relatively small number of total possible itineraries (28 distinct itineraries). From each itinerary, various data attributes are provided, including departure time, level of service (nonstop, single connection, double connection), carrier, fare ratio (the comparative fare levels, on average, across the airlines serving this city pair), and distance ratio (the ratio of itinerary flight distance to straight line distance). The data on the itineraries are shown in Table 5.1.

Table 5.1 Flight itinerary choices in synthetic data

Itinerary Number	Airline	Departure Time	Distance Ratio	Fare Ratio	Level of Service
1	BB	12:55	100	104	Non-stop
2	BB	21:05	100	104	Non-stop
3	AA	13:19	111	100	Single-Connect
4	AA	16:47	111	100	Single-Connect
5	AA	16:47	111	100	Single-Connect
6	AA	8:20	111	100	Single-Connect
7	AA	16:15	111	100	Single-Connect
8	CC	18:20	127	55	Single-Connect
9	CC	9:15	127	55	Single-Connect
10	BB	16:45	132	104	Single-Connect
11	BB	14:50	132	104	Single-Connect
12	BB	7:20	132	104	Single-Connect
13	BB	12:30	111	104	Single-Connect
14	BB	17:05	111	104	Single-Connect
15	BB	18:50	111	104	Single-Connect
16	BB	7:45	111	104	Single-Connect
17	DD	9:15	127	46	Single-Connect
18	DD	18:20	127	46	Single-Connect
19	CC	8:00	130	55	Single-Connect
20	BB	9:00	132	104	Single-Connect
21	AA	10:05	132	100	Double-Connect
22	AA	16:15	132	100	Double-Connect
23	AA	14:40	132	100	Double-Connect
24	BB	11:00	153	104	Double-Connect
25	DD	7:15	130	46	Double-Connect
26	DD	14:40	130	46	Double-Connect
27	EE	7:30	121	49	Double-Connect
28	EE	7:30	121	49	Double-Connect

The advantage of the HeNGEV model described in this chapter is that it can incorporate attributes of the decision-maker (or of the choice itself) into the correlation structure. To examine the usefulness of such enhanced tools, the dataset also includes information on the annual income level of each decision-maker, as well as the number of days in advance that the ticket was purchased.

The structure of this model is depicted in Figure 5.7. The network depicted has numerous nodes and arcs. If the associated parameters were each estimated

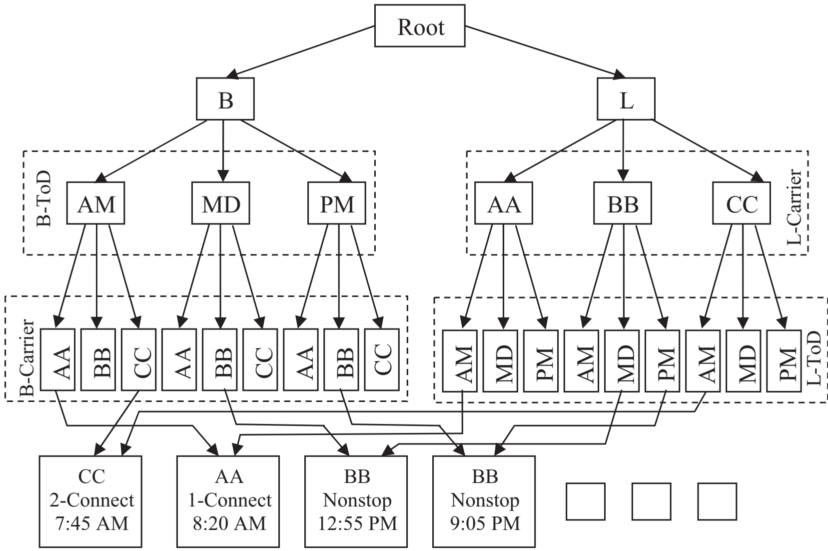


Figure 5.7 Flight itinerary choice model for synthetic data

independently, the parameter estimation process would become overwhelmed, and the resulting model would be virtually meaningless as a descriptive or predictive tool. Instead, the nodes are grouped into four sections (upper and lower nests on each side) with common logsum parameters, and the allocations between the sides are grouped together so that all alternatives would have common allocation parameters.

Since the data in this example are synthetic, the true model underlying the observations is known. In particular, the distribution of the covariance structure in the population is known and defined to be heterogeneous. This distribution is shown in Figure 5.8. A large share of the population is grouped near the right side, having a covariance structure nearly entirely defined by the *L* sub-model, whereas a much smaller share of the population is represented on the *B* sub-model side. This reflects the common scenario in air travel, where there are a few (generally high-revenue and business-related) travelers, who make decisions in a different way than most other travelers.

Estimated Models

The estimated parameters for the HeNGEV model are shown in Table 5.2. Most of the parameters in this model closely match the “true” parameters, although three, with bolded *t*-statistics, show a statistically significant difference from the true values. The fact that these three parameters are not correctly finding their true values is explained in part by the high correlation in their estimators, highlighted in Table 5.3.

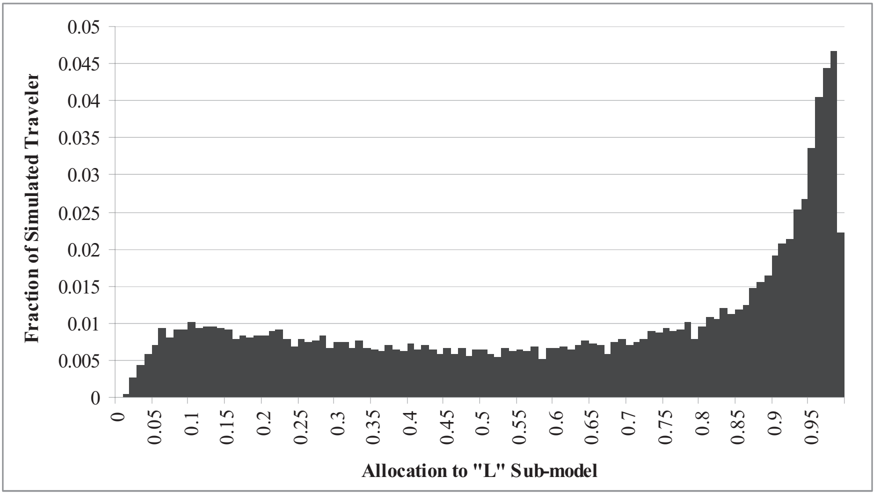


Figure 5.8 Distribution of allocation weights in unimodal synthetic data

Table 5.2 HeNGEV model

	True Value	Parameter Estimate	Std. Error of Estimate	t-stat vs. true
Departure Time				
Before 8 AM (ref.)	0	0	--	--
8–9:59 AM	0.15	0.0165	0.01796	**2.42**
10 AM–12:59 PM	0.10	0.09257	0.09851	0.08
1–3:59 PM	0.05	0.02468	0.02453	1.03
4–6:59 PM	0.10	0.07013	0.01876	1.60
7 PM or later	-0.30	-0.2975	0.09828	0.03
Level of Service				
Non-stop (ref.)	0	0	--	--
Single-connect	-2.3	-2.286	0.1019	0.14
Double-connect	-5.8	-5.864	0.1354	0.47
Flight Characteristics				
Distance Ratio	-0.01	-0.007141	0.001107	**2.58**
Fare Ratio	-0.004	-0.003359	0.0005518	1.16

Table 5.2 *Concluded*

	True Value	Parameter Estimate	Std. Error of Estimate	t-stat vs. true
Nesting Parameters				
B Time of Day (Upper) Nest	0.8	0.7994	0.01509	0.04
B Carrier (Lower) Nest	0.2	0.1439	0.02585	**2.17**
L Carrier (Upper) Nest	0.7	0.6746	0.01973	1.29
L Time of Day (Lower) Nest	0.3	0.3075	0.006947	1.08
Allocation Parameters				
Phi Constant L Side	1	1.066	0.3890	0.17
Phi Income (000) L Side	-0.03	-0.02912	0.005029	0.17
Phi Advance Purchase L Side	0.2	0.1772	0.02686	0.85
Model Fit Statistics				
LL at zero	-333220.45			
LL at convergence	-176880.64			
Rho-square w.r.t. zero	0.469			

The NetGEV model without a heterogeneous covariance (shown in Table 5.4) performs relatively well, but definitely worse than the HeNGEV model. The NetGEV model has a log likelihood at convergence that is 240 units smaller than the HeNGEV model, a highly significant deterioration given that only two degrees of freedom are lost. The performance of the individual parameter estimates in the NetGEV and HeNGEV models are compared in Table 5.5. For each parameter in the model, the HeNGEV estimate is closer to the known true value than the NetGEV estimate, generally by about half. Further, the standard errors of the estimates are all smaller for the HeNGEV model, also by about half.

For a more complete picture, regular NL models were estimated using each of the two sub-models, as well as a multinomial logit model that ignored the error covariance entirely. The results of these models are shown in Table 5.6. A graphical representation of the relationship between the various estimated models is shown in Figure 5.9. Not surprisingly, the MNL model with similarly defined

Table 5.3 Parameter estimator correlation, HeNGEV model

	08:00-09:59	10:00-12:59	13:00-15:59	16:00-18:59	19:00 or later	Distance Ratio	Fare Ratio	Single Connect	Double Connect	B Carrier (Lower) Nest	B Time of Day (Upper) Nest	L Time of Day (Lower) Nest	L Carrier (Upper) Nest	Phi Advance Purchase L Side	Phi Constant L Side	Phi Income (000) L Side
08:00-09:59	1.000	0.075	0.609	0.769	0.027	-0.901	-0.783	-0.124	-0.113	0.817	0.327	0.428	0.656	0.463	0.145	-0.411
10:00-12:59	0.075	1.000	0.052	0.132	0.996	-0.049	-0.026	0.958	0.737	0.061	-0.030	-0.317	0.059	0.022	0.004	-0.023
13:00-15:59	0.609	0.052	1.000	0.714	0.029	-0.547	-0.542	-0.050	0.006	0.561	-0.118	0.289	0.214	0.028	-0.075	-0.061
16:00-18:59	0.769	0.132	0.714	1.000	0.100	-0.661	-0.567	0.000	0.064	0.628	-0.016	0.216	0.354	0.132	-0.044	-0.150
19:00 or later	0.027	0.996	0.029	0.100	1.000	0.001	0.029	0.972	0.754	0.017	-0.070	-0.348	-0.007	-0.023	-0.011	0.016
Distance Ratio	-0.901	-0.049	-0.547	-0.661	0.001	1.000	0.870	0.133	0.141	-0.901	-0.336	-0.460	-0.685	-0.494	-0.168	0.439
Fare Ratio	-0.783	-0.026	-0.542	-0.567	0.029	0.870	1.000	0.198	0.220	-0.821	-0.409	-0.566	-0.723	-0.516	-0.155	0.461
Single-Connect	-0.124	0.958	-0.050	0.000	0.972	0.133	0.198	1.000	0.800	-0.110	-0.230	-0.466	-0.185	-0.178	-0.075	0.146
Double-Connect	-0.113	0.737	0.006	0.064	0.754	0.141	0.220	0.800	1.000	-0.112	-0.321	-0.419	-0.260	-0.265	-0.133	0.212
B Carrier (Lower) Nest	0.817	0.061	0.561	0.628	0.017	-0.901	-0.821	-0.110	-0.112	1.000	0.264	0.409	0.592	0.437	0.136	-0.398
B Time of Day (Upper) Nest	0.327	-0.030	-0.118	-0.016	-0.070	-0.336	-0.409	-0.230	-0.321	0.264	1.000	0.444	0.571	0.699	0.290	-0.598
L Time of Day (Lower) Nest	0.428	-0.317	0.289	0.216	-0.348	-0.460	-0.566	-0.466	-0.419	0.409	0.444	1.000	0.395	0.338	0.086	-0.304
L Carrier (Upper) Nest	0.656	0.059	0.214	0.354	-0.007	-0.685	-0.723	-0.185	-0.260	0.592	0.571	0.395	1.000	0.736	0.330	-0.598
Phi Advance Purchase L Side	0.463	0.022	0.028	0.132	-0.023	-0.494	-0.516	-0.178	-0.265	0.437	0.699	0.338	0.736	1.000	0.244	-0.702
Phi Constant L Side	0.145	0.004	-0.075	-0.044	-0.011	-0.168	-0.155	-0.075	-0.133	0.136	0.290	0.086	0.330	0.244	1.000	-0.811
Phi Income (000) L Side	-0.411	-0.023	-0.061	-0.150	0.016	0.439	0.461	0.146	0.212	-0.398	-0.598	-0.304	-0.598	-0.702	-0.811	1.000

Table 5.4 NetGEV model

	True Value	Parameter Estimate	Std. Error of Estimate	t-stat vs. true
Departure Time				
Before 8 AM (ref.)	0	0	--	--
8–9:59 AM	0.15	0.06687	0.03759	**2.21**
10 AM–12:59 PM	0.10	0.03704	0.1177	0.53
1–3:59 PM	0.05	-0.03495	0.07088	1.20
4–6:59 PM	0.10	0.02141	0.05334	1.47
7 PM or later	-0.30	-0.3445	0.1120	0.40
Level of Service				
Non-stop (ref.)	0	0	--	--
Single-connect	-2.3	-2.331	0.1407	0.22
Double-connect	-5.8	-5.956	0.2530	0.62
Flight Characteristics				
Distance Ratio	-0.01	-0.004372	0.002449	**2.30**
Fare Ratio	-0.004	-0.002202	0.001068	1.68
Nesting Parameters				
B Time of Day (Upper) Nest	0.8	0.8307	0.1022	0.30
B Carrier (Lower) Nest	0.2	0.07244	0.04395	**2.90**
L Carrier (Upper) Nest	0.7	0.6519	0.08702	0.55
L Time of Day (Lower) Nest	0.3	0.3078	0.01321	0.59
Allocation Parameters				
Phi Constant L Side	1	0.5928	0.4722	-0.86
Model Fit Statistics				
LL at zero	-333220.45			
LL at convergence	-177121.27			
Rho-square w.r.t. zero	0.468			

Table 5.5 Comparison of HeNGEV and NetGEV models

	HeNGEV Model		NetGEV Model	
	Actual Error of Estimate	**Std. Error of Estimate**	**Actual Error of Estimate**	**Std. Error of Estimate**
Departure Time				
Before 8 A.M. (ref.)	--	--	--	--
8–9:59 A.M.	-0.0435	0.01796	-0.08313	0.03759
10 A.M.–12:59 P.M.	-0.00743	0.09851	-0.06296	0.1177
1–3:59 P.M.	-0.02532	0.02453	-0.08495	0.07088
4–6:59 P.M.	-0.02987	0.01876	-0.07859	0.05334
7 P.M. or later	0.0025	0.09828	-0.0445	0.1120
Level of Service				
Non-stop (ref.)	0	0	--	--
Single-connect	0.014	0.1019	-0.031	0.1407
Double-connect	-0.064	0.1354	-0.156	0.2530
Flight Characteristics				
Distance Ratio	0.002859	0.001107	0.005628	0.002449
Fare Ratio	0.000641	0.0005518	0.001798	0.001068
Nesting Parameters				
B Time of Day (Upper) Nest	-0.0006	0.01509	0.0307	0.1022
B Carrier (Lower) Nest	-0.0561	0.02585	-0.1276	0.04395
L Carrier (Upper) Nest	-0.0254	0.01973	-0.0481	0.08702
L Time of Day (Lower) Nest	0.0075	0.006947	0.0078	0.01321
Allocation Parameters				
Phi Constant L Side	0.066	0.3890	-0.4072	0.4722
Phi Income (000) L Side	0.00088	0.005029		
Phi Advance Purchase L Side	-0.0228	0.02686		

utility functions performs relatively poorly, with log likelihood benefits in the thousands for a change to either nested structure.

The *L*-only structure has a better fit for the data than the *B*-only model. This is consistent with the construction of this dataset, which is heavily weighted with decision-makers exhibiting error correlation structures that are nearly the same as the *L*-only model. This heavy weight towards the *L* model is also reflected in the very small improvement (6.77) in log likelihood when moving from the

Table 5.6 Summary of model estimations

		HeNGEV Model		NetGEV Model		NL (L) Model		NL (B) Model		MNL Model	
	True Value	Estimated Parameter	Std. Err of Estimate	Estimated Parameter	Std. Error of Estimate	Estimated Parameter	Std. Error of Estimate	Estimated Parameter	Std. Error of Estimate	Estimated Parameter	Std. Error of Estimate
Departure Time											
Before 8 A.M. (ref)	0	0	--	0	--	0	--	0	--	0	--
8 – 9:59 A.M.	0.15	0.1065	0.01796	0.06687	0.03759	0.1615	0.01734	0.8323	0.1141	0.2668	0.02379
10 A.M. – 12:59 P.M.	0.10	0.09257	0.09851	0.03704	0.1177	0.09445	0.1003	-1.326	0.4197	-4.684	0.2893
1 – 3:59 P.M.	0.05	0.02468	0.02453	-0.03495	0.07088	-0.0211	0.02391	-1.303	0.3231	0.406	0.02834
4 – 6:59 P.M.	0.10	0.07013	0.01867	0.02141	0.05334	0.04509	0.01896	-0.8219	0.2305	-0.1938	0.02282
7 P.M. or later	-0.30	-0.2975	0.09828	-0.3445	0.1120	-0.3276	0.1001	-2.253	0.4913	-5.20	0.2894
Level of Service											
Non-stop (ref.)	0	0	--	0	--	0	--	0	--	0	--
Single-connect	-2.3	-2.286	0.1019	-2.331	0.1407	-2.455	0.1019	-6.552	0.8812	-7.355	0.289
Double-connect	-5.8	-5.864	0.1354	-5.956	0.2530	-6.274	0.1324	-16.19	2.098	-12.21	0.3015
Flight Characteristics											
Distance Ratio	-0.01	-0.00714	0.00111	-0.004372	0.00245	-0.01117	0.00081	-0.04809	0.00646	-0.07936	0.00136
Fare Ratio	-0.004	-0.00336	0.00055	-0.002202	0.00107	-0.00517	0.00045	-0.02619	0.00346	-0.03957	0.00046
Nesting Parameters											
B TOD (UN)	0.8	0.7994	0.01509	0.8307	0.1022			2.447	0.3128		
B Carrier (LN)	0.2	0.1439	0,02585	0.07244	0.04395			0.8607	0.1110		
L Carrier (UN)	0.7	0.6746	0.01973	0.6519	0.08702	0.8193	0.01063				
L TOD (LN)	0.3	0.3075	0.00695	0.3078	0.01321	0.3133	0.0061				
Allocation Parameters (L Side)											
Phi Constant	1	1.066	0.389	0.5928	0.4722						
Phi Income (000)	-0.03	-0.0291	0.00503								
Phi Adv. Pur.	0.2	0.1772	0.02686								
Model Fit Statistics											
LL at zero		-333220		-333220		-333220		-333220		-333220	
LL at convergence		-176881		-177121		-177128		-177244		-180964	
Rho-square w.r.t. zero		0.469		0.468		0.468		0.468		0.457	

Key: TOD = Time of Day; UN = Upper Nest; LN = Lower Nest

Source: Adapted from Newman 2008a: Table 6.5 (reproduced with permission of author).

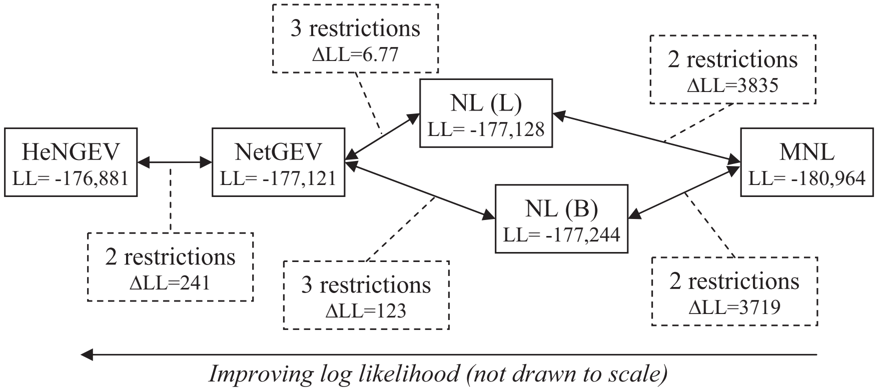


Figure 5.9 Log likelihoods and relationships among models estimated using unimodal dataset

L-only model to the NetGEV model, which incorporates both *L*- and *B*-sub-models. Although this change is still statistically significant ($\chi^2 = 13.54$, with three degrees of freedom, $p = 0.0036$) it is small compared to the changes observed between other models. In this instance, with most travelers exhibiting similar *L*-choice patterns, it appears that upgrading to the NetGEV model alone does not provide much benefit. Far more improvement in the log likelihood is made when the heterogeneous covariance is introduced, which allows the small portion of the population that exhibits "*B*" choice patterns to follow that model, without adversely affecting the predictions for the larger *L*-population.

The predictions of the HeNGEV model and the NetGEV model across the entire market are roughly similar, as can be seen in Table 5.7. The two models over- or under-predict in roughly the same amounts for each itinerary. However, when the predictions are segmented by income as in Table 5.8, the HeNGEV model can be seen to outperform the NetGEV model in all income segments, especially in the extremes of the income range. The errors for the whole market, on the right side of Figure 5.10, are roughly similar for both models. However, within the extreme high and low income segments (especially in the high income segment), as shown in Figure 5.11, the errors in prediction for the HeNGEV model are generally much smaller than those of the NetGEV model. The overall market predictions for the NetGEV model end up close to the HeNGEV predictions because the particularly large errors appearing in the extreme income segments have offsetting signs.

Discussion

Overall, the HeNGEV models show a better fit for the synthetic data than the matching homogeneous NetGEV models. The HeNGEV models give significantly better log likelihoods in both the bimodal and unimodal scenarios, indicating that

Table 5.7 HeNGEV and NetGEV market-level predictions

		Predictions		Differences	
Itinerary	Total Observed	HeNGEV	NetGEV	HeNGEV	NetGEV
1	45067	44806.47	44824.55	-260.53	-242.45
2	26746	26769.61	26753.70	23.61	7.70
3	2633	2649.82	2650.90	16.82	17.90
4	1346	1439.44	1432.45	93.44	86.45
5	1415	1439.44	1432.45	24.44	17.45
6	3521	3328.98	3355.50	-192.02	-165.50
7	1452	1439.44	1432.45	-12.56	-19.55
8	3328	3273.62	3293.55	-54.38	-34.45
9	2374	2485.81	2466.85	111.81	92.85
10	13	13.63	16.25	0.63	3.25
11	4	5.91	7.050	1.91	3.05
12	432	481.71	480.35	49.71	48.35
13	10	12.00	12.00	2.00	2.00
14	24	22.22	21.90	-1.78	-2.10
15	20	22.22	21.90	2.22	1.90
16	1047	1055.51	1053.15	8.51	6.15
17	3983	4014.62	4001.65	31.62	18.65
18	3412	3506.99	3506.00	94.99	94.00
19	2221	2,257.96	2264.90	36.96	43.90
20	819	834.07	831.55	15.07	12.55
21	0	0.00	0.00	0.00	0.00
22	0	0.00	0.00	0.00	0.00
23	0	0.00	0.00	0.00	0.00
24	0	0.00	0.00	0.00	0.00
25	1	0.00	0.00	-1.00	-1.00
26	16	21.71	20.65	5.71	4.65
27	61	59.41	60.15	-1.59	-0.85
28	55	59.41	60.15	4.41	5.15

Table 5.8 HeNGEV and NetGEV predictions segmented by income

	Observed Choices					HeNGEV Model					NetGEV Model				
Itin	Bottom Fifth		Middle Fifth		Top Fifth	Bottom Fifth		Middle Fifth		Top Fifth	Bottom Fifth		Middle Fifth		Top Fifth
1	8884	8958	9010	9139	9076	11.5	-27.3	-53.6	-152.0	-39.2	80.9	6.9	-45.1	-174.1	-111.1
2	5246	5211	5264	5423	5602	-128.2	33.0	72.5	23.3	23.0	104.7	139.7	86.7	-72.3	-251.3
3	572	565	533	500	463	-6.4	-18.5	-0.4	16.0	26.1	-41.8	-34.8	-2.8	30.2	67.2
4	275	285	280	277	229	48.0	19.2	10.5	-2.9	18.6	11.5	1.5	6.5	9.5	57.5
5	292	332	261	285	245	31.0	-27.8	29.5	-10.9	2.6	-5.5	-45.5	25.5	1.5	41.5
6	703	730	722	686	680	-37.0	-64.1	-56.2	-20.3	-14.4	-31.9	-58.9	-50.9	-14.9	-8.9
7	307	318	292	260	275	16.0	-13.8	-1.5	14.2	-27.4	-20.5	-31.5	-5.5	26.5	11.5
8	693	730	681	622	602	16.7	-49.7	-22.2	11.2	-10.4	-34.3	-71.3	-22.3	36.7	56.7
9	503	495	497	460	419	26.1	17.1	2.5	24.7	41.5	-9.6	-1.6	-3.6	33.4	74.4
10	6	3	0	1	3	-2.2	0.2	2.8	1.3	-1.6	-2.8	0.3	3.3	2.3	0.3
11	2	1	0	0	1	-0.3	0.4	1.2	1.0	-0.4	-0.6	0.4	1.4	1.4	0.4
12	78	78	84	95	97	12.7	15.7	11.9	3.6	5.8	18.1	18.1	12.1	1.1	-0.9
13	5	1	2	1	1	-1.6	1.9	0.5	1.0	0.3	-2.6	1.4	0.4	1.4	1.4
14	9	6	2	5	2	-2.8	-0.7	2.6	-1.3	0.4	-4.6	-1.6	2.4	-0.6	2.4
15	5	7	2	3	3	1.3	-1.7	2.6	0.7	-0.6	-0.6	-2.6	2.4	1.4	1.4
16	181	181	228	226	231	-11.2	10.9	-20.0	1.3	27.5	29.6	29.6	-17.4	-15.4	-20.4
17	842	803	822	761	755	-6.8	14.9	-16.7	29.3	10.9	-41.7	-2.7	-21.7	39.3	45.3
18	740	675	715	625	657	27.2	57.0	-8.7	50.7	-31.1	-38.8	26.2	-13.8	76.2	44.2
19	477	462	416	442	424	11.6	6.8	38.3	-4.9	-14.9	-24.0	-9.0	37.0	11.0	29.0
20	148	134	164	159	214	-7.3	20.7	0.9	18.0	-17.2	18.3	32.3	2.3	7.3	-47.7
21	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
22	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
23	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
24	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25	0	0	1	0	0	0.0	0.0	-1.0	0.0	0.0	0.0	0.0	-1.0	0.0	0.0
26	2	2	5	1	6	1.9	2.2	-0.7	3.5	-1.2	2.1	2.1	-0.9	3.1	-1.9
27	15	12	10	13	11	0.0	1.3	2.1	-2.3	-2.7	-3.0	0.0	2.0	-1.0	1.0
28	15	11	9	16	4	0.0	2.3	3.1	-5.3	4.3	-3.0	1.0	3.0	-4.0	8.0
			Total Absolute Deviation:			407.87	407.2	361.9	399.51	321.9	530.6	519.19	369.9	564.37	884.23

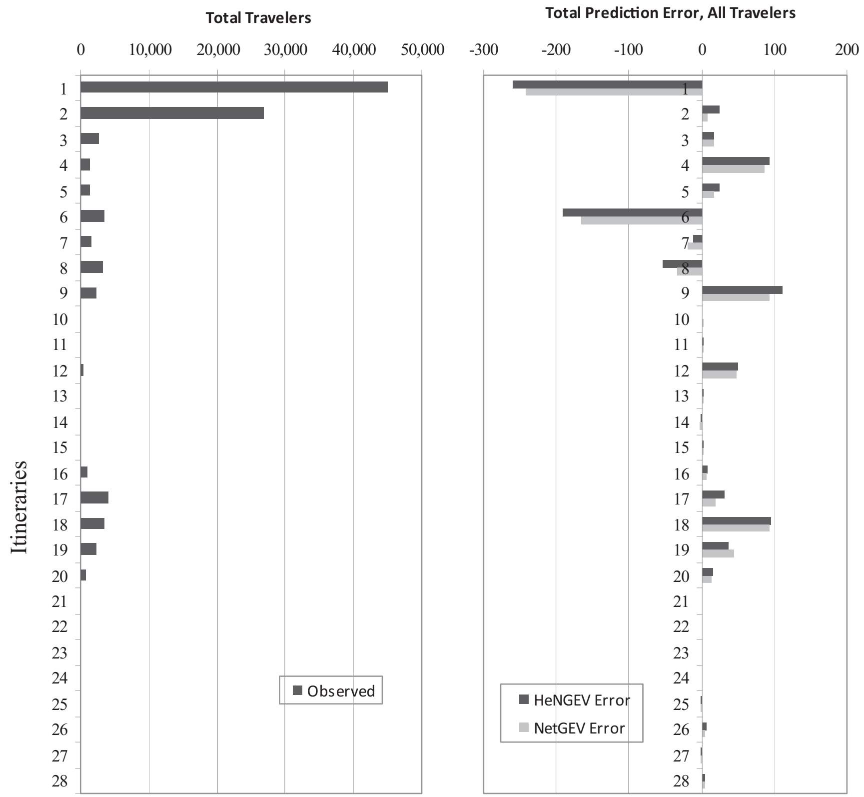


Figure 5.10 Observations and market-level prediction errors

this model type may be useful in a variety of situations, even when the fraction of the population exhibiting "unusual" behavior is small. Individual parameter estimates were generally improved by adopting the heterogeneous model, often by half or more of the error in the estimate.

Better fitting models are obviously a positive attribute of the HeNGEV structure, but they are not the only benefit. When used to predict choices of subsections of the population, the responsiveness of the correlation structure to data allows the HeNGEV to be a superior predictive tool. Such benefits could be especially appealing in revenue management systems, which seek specifically to segment markets in order to capture these types of differences in pricing and availability decisions.

Summary of Main Concepts

This chapter presented an overview of the Network GEV (NetGEV) model. The NetGEV is a GEV model that contains at least three (and possibly more) levels.

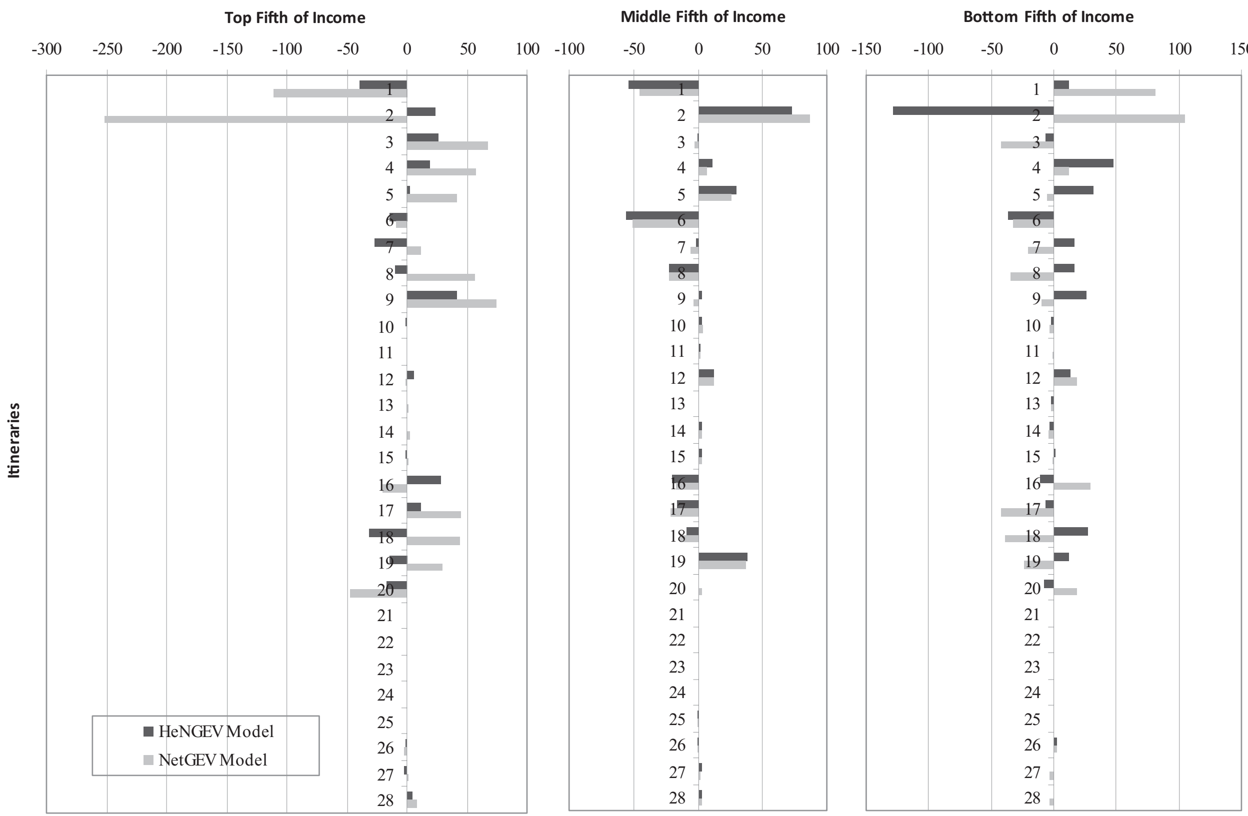


Figure 5.11 Prediction errors, segmented by income

The GNL model, which is a GEV model with two levels, is a special case of the NetGEV model. The NetGEV model is a relatively recent addition to the literature and provides a theoretical foundation for investigating properties of the hybrid, multi-level itinerary choice models proposed by Koppelman and Coldren (2005a, 2005b) that were introduced in Chapter 4.

The most important concepts covered in this chapter include the following:

- Normalizations are required when a model is over-specified, i.e., there is not a unique solution.
- The normalization rules presented in this chapter are just one of many possible normalization rules. For example, in a network that is both crash free and crash safe, either set of normalization rules may be applied and will result in unbiased parameter estimates.
- The network structure itself may lead to over-specification. In this case, the analyst needs to change the network structure, which in turn will result in a different covariance matrix, different choice model, and potentially different choice probabilities.
- Similar to the NL or GNL model, the logsum parameters in a NetGEV model are over-specified. It is common to normalize that logsum of the root node to one, which results in the familiar bounds of $0 < \mu_n \leq 1$. In addition, the logsum parameters associated with predecessor nodes (or nests higher in the tree) must be larger than the logsum parameters of successor nodes (or nests lower in the tree) to maintain positive covariance (and increased substitution) among alternatives that share a common nest.
- Although the normalization of logsum parameters in a NetGEV model is straightforward, normalization of allocation parameters is more involved. Fundamentally, this is due to the need to properly account for inter-elemental covariance when pieces of an elemental alternative are recombined prior to the root node.
- A crash free network is one in which multiple pieces of the same alternative are recombined only at the root node. In this case, setting the NetGEV allocation terms a_{ij} to the familiar allocation weights presented in Chapter 4 (the τ_{ij}'s) is a valid normalization. In a crash free network, partial alternatives are recombined at the root node and no crashes occur, as there is no opportunity for internal correlation at intermediate nodes.
- A crash safe network is one in which only elemental alternative nodes have multiple predecessor nodes. In this case, a normalization is possible that effectively rescales the partial alternatives when they are recombined at an intermediate node. This normalization accounts for inter-elemental covariance, i.e., although there is the potential for a crash as alternatives recombine at an intermediate node, the crash can be avoided through appropriate rescaling of the allocation parameters.
- Heterogeneity in decision-maker preferences can be accommodated in a NetGEV model by allowing the allocation parameters to be a function of

observable decision-maker or trip-making characteristics. The resulting Heterogeneous Network GEV model (HeNGEV) may be particularly relevant in the airline applications, due to the fundamental differences between business and leisure passengers.

- Understanding the properties of the NetGEV model and determining how it is related to other known models in the literature is still a very active area of research. From a practical point of view, though, it is important to note that the primary motivation for using NetGEV models is to incorporate more realistic substitution patterns across alternatives. Often, these substitution patterns correspond to a well-defined network structure. All of the GEV models presented in Chapter 4, for instance, exhibit both the crash free and crash safe network properties. In this context, although the NetGEV is a very flexible model (and interesting to explore in a theoretical context), those network models motivated from a behavioral perspective will be straight-forward to normalize, estimate, and interpret.

Appendix 5.1: Nonlinear Constrained Splitting

If the structure of the GEV network conforms to neither crash free nor crash safe forms, and it is undesirable to include a full set of alternative specific constants, it may still be possible to build an unbiased model through constraints on the form of the allocation values, although these constraints will typically be complex and nonlinear. This appendix provides an example of one normalization procedure (which is much more complex than the crash free and crash safe normalizations presented earlier). The easiest way to find the necessary constraints is to decompose the network so that it has the structure needed to apply the crash safe normalizations.

For any network node $i \in N$ that has more than one incoming edge (i.e., $\left|i^{\uparrow}\right| = z > 1$), the network can be restructured by replacing i with z new nodes $i_1, i_2, \ldots, i_z$, each of which has the same μ value and the same set of outgoing edges to successor nodes, but only a single incoming edge from a single predecessor node: $j_1 \rightarrow i_1, j_2 \rightarrow i_2, \ldots, j_z \rightarrow i_z$. For each successor node k, the incoming edge from i is replaced with z new incoming edges from $i_1, i_2, \ldots, i_z$. Setting $a_{i_n k_n} = a_{j_n i} a_{i k_n}$ and $a_{j_n i_n} = 1$ for all $n \in \{1,2,\ldots,z\}$ will ensure that all nodes in the model excluding i will maintain the same G values, therefore preserving the model probabilities exactly.

This can be applied recursively through the network to split any nesting node which has multiple incoming edges. Since G is circuit free, and the splitting process can only increase the number of incoming edges on successor nodes, the entire network can be restructured to the desired form in a finite number of steps. In each node split, the number of edge allocation values is increased (more edges are added than removed), but the relationship between the allocation values of the additional edges is such that the number of values that can be independently determined remains constant. The final network can then be normalized according to the crash safe algorithm, subject to the constraints developed in the network decomposition process. A simple network is illustrative of the decomposition process as well as the potential complexity of the nonlinear constraints.

For example, consider the simple network depicted in Figure 5.12, which has two elemental alternative nodes, A and B, a root node R, and two other intermediate nesting nodes, H and L. This network conforms to neither the crash free form ($R \rightarrow H \rightarrow L \rightarrow B$ and $R \rightarrow H \rightarrow B$ diverge from each other at H, but diverge from $R \rightarrow L \rightarrow B$ at R) nor the crash safe form ($R \rightarrow H \rightarrow L \rightarrow B$ and $R \rightarrow L \rightarrow B$ converge at L, before converging with $R \rightarrow H \rightarrow B$ at B).

The network can be decomposed by splitting L into two new nodes, M and N. One of these nodes inherits the incoming edge from R, whereas the other inherits the incoming edge from H. Both M and N retain outbound edges to both A and B. The revised network is shown in Figure 5.13.

Unlike the original network in Figure 5.12, the revised network has some constraints imposed on its parameters:

$$\mu_M = \mu_N$$

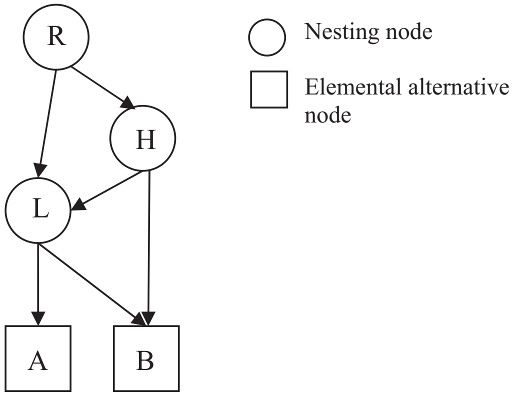


Figure 5.12 A simple network which is neither crash free nor crash safe

Source: Adapted from Newman 2008b: Figure 4 (reproduced with permission of Elsevier).

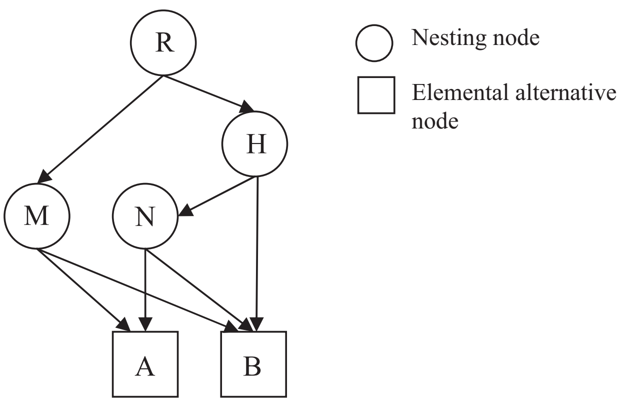


Figure 5.13 A revised network which is crash safe

Source: Adapted from Newman 2008b: Figure 5 (reproduced with permission of Elsevier).

$$a_{HN} = 1$$

$$a_{RM} = 1$$

$$a_{MA} / a_{NA} = a_{MB} / a_{NB} \tag{5.6}$$

The ratio constraint in Equation 5.6 arises from the replacement of a single allocative split at L in Figure 5.12 with two such splits, at M and N, in Figure 5.13. These two splits need to have the same relative ratio, as they are both “controlled” by the ratio of the single split in the original network.

The revised network now meets the structural requirements for crash safe normalization, as only nodes A and B have more than one incoming edge. This normalization replaces the a values with the new values:

$$a_{HB} = \left(\frac{\alpha_{HB}}{\alpha_{HB} + \alpha_{NB}} \right)^{\mu_H} \left(\alpha_{HB} + \alpha_{NB} \right)^{\mu_R}$$

$$a_{NB} = \left(\frac{\alpha_{NB}}{\alpha_{HB} + \alpha_{NB}} \right)^{\mu_H} \left(\alpha_{HB} + \alpha_{NB} \right)^{\mu_R}$$

$$a_{MB} = \left(1 - \alpha_{HB} - \alpha_{NB}\right)^{\mu_R}$$

$$a_{MA} = \alpha_{MA}^{\mu_R}$$

$$a_{NA} = \left(1 - \alpha_{NA}\right)^{\mu_R}$$

But from Equation 5.6:

$$\alpha_{MA} = \left(\frac{\left(\alpha_{NB}\right)^{\mu_H/\mu_R} \left(\alpha_{HB} + \alpha_{NB}\right)^{1-(\mu_H/\mu_R)}}{1 - \left(\alpha_{HB} + \alpha_{NB}\right)} + 1 \right)^{-1}$$

which is clearly a nonlinear constraint when $0 < \mu_H < \mu_R$.

The shape of the constraint for various different values of μ_H / μ_R is depicted in Figure 5.14. Each constraint surface is depicted inside a unit cube, as each α parameter must fall inside the unit interval, and each surface is defined exclusively in the left triangular region of the cube, because $\alpha_{HB} = \alpha_{NB} \leq 1$. In the upper left cube, where $\mu_H / \mu_R = 1$, the contour lines of constant α_{MA} are straight, as in that

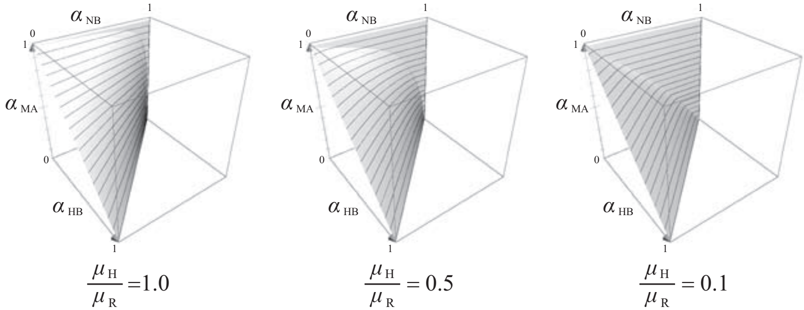


Figure 5.14 Constraint functions for various ratios of μ_H and μ_R

Source: Adapted from Newman 2008b: Figure 6 (reproduced with permission of Elsevier).

scenario α_{HB} and α_{NB} are linearly related when α_{MA} is otherwise fixed. As μ_H / μ_R approaches 0, the surface of the constraint asymptotically approaches the limiting planes of $\alpha_{MA} + \alpha_{HB} + \alpha_{NB} = 1$ and $\alpha_{HB} = 0$.

This page has been left blank intentionally

Chapter 6
Mixed Logit

Introduction

Chapter 5 portrayed the historical development of choice models as one that evolved along two research paths. On the surface, these paths appear to be quite distinct. The first focused on incorporating more flexible substitution patterns and correlation structures while maintaining closed-form expressions for the choice probabilities, resulting in the development of models that belong to the GNL and/or NetGEV class. The second focused on reducing computational requirements associated with numerically evaluating the likelihood function for the probit model.

In the late 1990's, however, advancements in simulation techniques enabled these two paths to converge, resulting in a powerful new model—the mixed logit—that has been shown to theoretically approximate *any* random utility model (Dalal and Klein 1988; McFadden and Train 2000). Like the probit, the mixed logit has a likelihood function that must be numerically evaluated. Distinct from the probit, however, numerical evaluation of integrals is facilitated by embedding the MNL probability as the "core" within the likelihood function. In this sense, the simplicity of the MNL probability is married with the complexity of integrals, the latter of which provide the ability to incorporate random taste variation, correlation across alternatives and/or observations, and/or heteroscedasticity.

To date, several aviation applications of mixed logit models have occurred. The majority of these applications have been done by the academic community using stated preference surveys or publically available datasets. There has been a very limited involvement of the aviation professional community in investigating the benefits of using these models to support revenue management, scheduling, marketing, and other critical business areas. The objective of this chapter is to present an overview of the mixed logit model, highlighting key concepts for researchers and practitioners venturing into this modeling area. For additional information, readers are referred to the textbook by Train (2003).

The next section provides an overview of initial mixed logit applications to both transportation (broadly defined) and aviation specifically. Next, two common formulations for the mixed logit model are presented: the random coefficients mixed logit and the error components mixed logit. Finally, identification rules for mixed logits, many of which evolved out of earlier work done in the context of probit models, are described. The chapter concludes with a summary of the main concepts.

History and Early Applications

Historically, the first applications of the mixed logit models occurred in the early 1980's by Boyd and Mellman (1980) and Cardell and Dunbar (1980). These early studies were based on aggregate market share data. Some of the first studies to use disaggregate individual or household data, including those of Train, McFadden, and Ben-Akiva (1987b), Chintagunta, Jain, and Vilcassim (1991), and Ben-Akiva, Bolduc, and Bradley (1993), used a quadrature technique to approximate one or two dimensions of integration. However, due to limitations of quadrature techniques for integrals of more than two dimensions, e.g., an inability to compute integrals with sufficient precision and speed for maximum likelihood applications (Hajivassiliou and Ruud 1994), it was not until simulation tools became more advanced that the mixed logit model became widely used.

Early applications of mixed logit models spanned individuals' residential and work location choices (e.g., Bolduc, Fortin and Fournier 1996; Rouwendahl and Meijer 2001), travelers' departure time, route, and mode choices (e.g., Cherchi and Ortuzar 2003; de Palma, Fontan and Picard 2003; Hensher and Greene 2003, etc.) consumers' choices among energy suppliers (e.g., Revelt and Train 1999), refrigerators (e.g., Revelt and Train 1998), automobiles (e.g., Brownstone Bunch and Train 2000), and fishing sites (e.g., Train 1998). The degree to which the discrete choice modeling community has embraced mixed logit models is evident in Table 6.1. The table synthesizes early mixed logit applications solved via numerical approximation simulation methods that appeared in the literature from 1996 to 2003. The table provides information on many of the concepts that will be discussed in this chapter including the type of application and data, i.e., revealed and/or stated preference; type of distribution(s) assumed and whether the distributions are independent or have a non-zero covariance; number of observations in the estimation dataset; number of fixed and random coefficients considered in the model specification(s); and number and types of draws used as support points. Studies based on simulated data and advanced mixed logit applications (e.g., ordered mixed logit or models that combine closed-form GEV and mixed logit applications) are excluded from the table but integrated throughout the discussion (see Bhat (2003a) for a review of these models). Also, although it would be interesting to compare the number of alternatives used in the empirical applications, few studies provided explicit information about the universal choice set alternatives; consequently, this information is excluded.

The number of publications using mixed logit models has expanded exponentially since 2003 and mixed logit models have been applied in numerous other transportation contexts spanning activity-based planning and rescheduling behavior models (Akar, Clifton and Doherty 2009; van Bladel, Bellemans, Janssens and Wets 2009; Bellemans, van Bladel, Janssens, Wets and Timmermans 2009), mode choice models (Duarte, Garcia, Limao and Polydoropoulou 2009; Meloni, Bez and Spissu 2009), residential location/relocation decisions (Eluru, Senar, Bhat, Pendyala and Axhausen 2009; Habib and Miller 2009), pedestrian

Table 6.1 Early applications of mixed logits based on simulation methods

Study	Application (Choice of...)[1]	Data	Distribution	Covariance included? (if yes, # of parameters)	# observations (# of individuals)[2]	# of fixed parameters	# of random parameters	# of draws	Type of draws
Bolduc, Fortin & Fournier (1996)	Doctor's office location	RP	Normal	Yes (NR)[3]	4369	22	5	50	NR[3]
Bhat (1998a)	Mode/dept time	RP	Normal	No	3000	5	6	500	NR
Bhat (1998b)	Mode	RP	Normal	No	2000	12	4	1000	NR
Revelt & Train (1998)	Refrigerator	Joint RP/SP	Normal, lognormal	Yes (all)	6081(410) SP; 163 RP	1	6	500	NR
"	Refrigerator	SP	Normal	No	375	6	6	500	NR
Train (1998)	Fishing site	RP	Normal, lognormal	Yes (3)	962 (259)	1	7	1000	NR
Brownstone & Train (1999)	Automobile	SP	Normal	No	4656	21	4	250	NR
Revelt & Train (1999)	Energy supplier	SP	Normal, lognormal uniform, triangular	No	4308 (361)	1	5	NR	Halton
Bhat (2000b)	Mode	RP	Normal	No	2806 (520)	5	7	1000	NR
Brownstone, Bunch & Train (2000)	Automobile	Joint RP/SP	Normal	No	4656 SP; 607 RP	16-28	5	1000[5]	NR
Goett, Hudson & Train (2000)	Energy supplier	SP	Normal	No	4820 (1205) per segment	2	9-15	250	Halton
Kawamura (2000)	Truck VOT	SP	Lognormal	No	350-985 (70)	0	2	NR	NR
Calfee, Winston & Stempski (2001)	Auto VOT	SP	Normal, lognormal	No	1170	2	2	100	Random
Han, Algers & Engelson (2001)	Route/VOT	SP	Normal, uniform	No	1157 (401)	0	9	1000[6]	Random
Hensher (2001a)	Route/VOT	SP	Normal, lognormal uniform, triangular	Yes[4]	3168 (198)	1-2	4-5	50	Halton

[1]Due to space considerations, the type of mode, route, or value of time (VOT) study is not further classified. [2]Number in parenthesis reflects the number of individuals providing multiple SP responses. [3]Not reported (abbreviated as NR). [4]Assumes a parametric form for unobserved spatial correlation based on distance function. [5]Draws increased to 1000 for numerical stability. [6]Authors tested 10 to 2000 draws and note appropriate number is application specific. [7]Authors tested 10 to 200 Halton draws and found 50 draws to produce stable VOT estimates. [8]30 SP choices per 264 individuals has been assumed. [9]Assumes a parametric covariance form proportional to a path attribute. [10]Instability in parameter estimates seen with 100,000 draws. [11]Draws increased from 1500 due to sensitivity in standard errors.

Source: Modified from Garrow 2004: Table 2.2 (reproduced with permission of author).

Table 6.1 *Concluded*

Study	Application (Choice of...)[1]	Data	Distribution	Covariance included? (if yes, # of parameters)	# observations (# of individuals)[2]	# of fixed parameters	# of random parameters	# of draws	Type of draws
Hensher (2001b)	Route/VOT	SP	Triangular	Yes (6)	2304 (144)	2	10	50[7]	Halton
Rouwendahl & Meijer (2001)	Residential & work location	SP	Normal	No	7920[8] (264)	1-16	21	250	NR
Beckor, Ben-Akiva & Ramming (2002)	Route	RP	Normal	Yes[9]	159	12	1	4069 to 100000[10]	NR
Small, Winston & Yan (2005) ; working paper in 2002	Route (toll)	Joint RP/SP	Normal	No	641 (82) SP; 82 RP	14	2	2000[11]	Random
"	Route (toll)	SP	Normal	No	641 (82) SP	4	3	1000	Random
Bhat & Gossen (2004); working paper in 2003	Weekend activity type	RP	Normal	Yes (all)	3493 (2390)	23	3	NR	Halton
Brownstone & Small (2003)	Route (toll)	SP	Not mentioned	No	601	6	3	NR	NR
Cherchi & Ortuzar (2003)	Mode	RP	Normal	No	338	10-14	1-2	NR	NR
de Palma, Fontan & Picard (2003)	Dept. time	RP	Lognormal	No	1941	2	2	10000	NR
"	Dept. time	RP	Lognormal	No	987	5	2	10000	NR
"	Dept. time	RP	Lognormal	No	835	6	1	10000	NR
Hensher & Greene (2003)	Route	SP	Lognormal	No	4384 (274)	7	1	25-2000	Halton
"	Route	SP	Lognormal	No	2288 (143)	7	1	25-2000	Halton
"	Route	RP	Lognormal	No	210	7	1	25-2000	Halton

[1]Due to space considerations, the type of mode, route, or VOT study is not further classified. [2]Number in parenthesis reflects the number of individuals providing multiple SP responses. [3]Not reported (abbreviated as NR). [4]Assumes a parametric form for unobserved spatial correlation based on distance function. [5]Draws increased to 1000 for numerical stability. [6]Authors tested 10 to 2000 draws and note appropriate number is application specific. [7]Authors tested 10 to 200 Halton draws and found 50 draws to produce stable VOT estimates. [8]30 SP choices per 264 individuals has been assumed. [9]Assumes a parametric covariance form proportional to a path attribute. [10]Instability in parameter estimates seen with 100,000 draws. [11]Draws increased from 1500 due to sensitivity in standard errors.

injury severity (Kim, Ulfarsson, Shankar and Mannering 2009), bicyclist behavior (Sener, Eluru and Bhat 2009), consideration of physical activity in choice of mode (Meloni, Portoghese, Bez and Spissu 2009), and response of automakers' vehicle designs due to regulations (Shiau, Michalek and Hendrickson 2009).

Applications of mixed logit models to aviation began to appear around 2003. As shown in Table 6.2, the majority of these earliest applications were based on stated preference surveys, often in the context of multiple airport choice (e.g., Hess and Polak 2005a, 2005b; Hess 2007; Pathomsiri and Haghani 2005), carrier/itinerary choice (e.g., Adler, Falzarano and Spitz 2005; Collins, Rose and Hess 2009; Warburg, Bhat and Adler 2006; Wen, Chen and Huang 2009) and intercity mode choice in which train, auto, and/or bus substitution with air was examined (e.g., Carlsson

Table 6.2 Aviation applications of mixed logit models

Study	Application	Data
Carlsson (2003)	Business travelers' intercity mode choice in Sweden (choice of rail/air)	SP
Garrow (2004)	Air travelers' show, no show, and day of departure standby behavior	RP data from a major US airline
Adler, Falzarano and Spitz (2005)	Itinerary choice with airline and access effects	SP
Hess and Polak (2005a)	Airport choice	SP
Hess and Polak (2005b)	Airport, airline, access choice	1995 San Francisco Air Passenger Survey (MTC 1995)
Pathomsiri and Haghani (2005)	Airport choice	SP
Lijesen (2006)	Value of flight frequency	SP
Srinivasan, Bhat and Holguin-Veras (2006)	Intercity mode choice (with 9/11 security effects)	SP
Warburg, Bhat and Adler (2006)	Business travelers' itinerary choice	SP
Ashiabor, Baik and Trani (2007)	Air/auto mode choice by U.S. county and commercial service airports (developed for NASA to predict demand for small aircraft)	1995 American Travel Survey (BTS 1995)

Table 6.2 Concluded

Study	Application	Data
Hess (2007)	Airport and airline choice	SP
Collins, Rose and Hess (2009)	Comparison of willingness to pay estimates between a traditional SP survey and a "mock" on-line travel agency survey	SP
Wen, Chen and Huang (2009)	Taiwanese passengers' choice of international air carriers (service attributes)	SP
Xu, Holguin-Veras and Bhat (2009)	Intercity mode choice (with airport screening time effects after 9/11)	SP
Yang and Sung (2010)	Introduction of high speed rail in Taiwan (competition with air, bus, train)	SP

Note: MTC = Metropolitan Transport Commission. BTS = Bureau of Transportation Statistics.

2003; Srinivasan, Bhat and Holguin-Veras 2006; Ashiabor, Baik and Trani 2007; Xu Holguin-Veras and Bhat 2009; Yang and Sung 2010). Another unique application included the use of stated preference surveys to examine how customers value flight frequency (Lijesen 2006). To the best of the author's knowledge, there have been no applications of mixed logit models based on proprietary airline datasets, aside from Garrow (2004) in the context of no show models.

Random Coefficients Interpretation for Mixed Logit Models

Two primary formulations or interpretations of mixed logit probabilities exist, which differ depending on whether the primary objective is to: 1) incorporate random taste variation; or, 2) incorporate correlation and/or unequal variance across alternatives or observations. These different objectives led to different names for the "mixed logit" models in early publications, before the term "mixed logit" was generally adopted by the discrete choice modeling community. That is, mixed logits have also been called random-coefficients logit or random-parameters logit (e.g., Bhat 1998b; Train 1998), error-components logit (e.g., Brownstone and Train 1999), logit kernel (e.g., Beckor, Ben-Akiva and Ramming 2002; Walker 2002), and continuous mixed logit (e.g., Ben-Akiva, Bolduc and Walker 2001).

Conceptually, the mixed logit model is identical to the MNL model except that the parameters of the utility functions for mixed models can vary across

individuals, alternatives, and/or observations. However, this added flexibility comes at a cost—choice probabilities can no longer be expressed in closed-form. Under a random parameters formulation, the utility that individual n obtains from alternative i is given as $U_{ni} = \boldsymbol{\beta}' x_{ni} = \varepsilon_{ni}$ where $\boldsymbol{\beta}$ is the vector of parameters associated with attributes x_{ni}, and ε_{ni} is a random error component. Unlike the MNL model, the $\boldsymbol{\beta}$ parameters are no longer fixed values that represent "average" population values, but rather are random realizations from the density function $f(\boldsymbol{\beta})$. Thus, mixed logit choice probabilities are expressed as the integral of logit probabilities evaluated over the density of distribution parameters, or

$$P_{ni} = \int L_{ni}(\beta) f(\beta \mid \eta) d\beta \tag{6.1}$$

where:

P_{ni} is the probability individual n chooses alternative i,

$L_{ni}(\boldsymbol{\beta})$ is a logit probability evaluated at the vector of parameter estimates $\boldsymbol{\beta}$ that are random realizations from the density function $f(\boldsymbol{\beta})$,

$\boldsymbol{\eta}$ is a vector of parameter estimates associated with the density function $f(\boldsymbol{\beta})$.

In a mixed model, L_{ni} takes the MNL form. For example, for a particular realization of $\boldsymbol{\beta}$, the mixed MNL logit probability is:

$$L_{ni}(\beta) = \frac{\exp(V_{ni})}{\sum\limits_{j \in C_n} \exp(V_{nj})}$$

where:

C_n is the set of alternatives available in the choice set for individual n.

The problem of interest is to solve for the vector of distribution parameters $\boldsymbol{\eta}$ associated with the $\boldsymbol{\beta}$ coefficients given a random sample of observations from the population. Distinct from the formulation of the GNL and NetGEV, some or all of the $\boldsymbol{\beta}$ coefficients are assumed to vary in an unspecified, therefore "random," pattern.

From a modeling perspective, the analyst begins with the assumption that individuals' "preferences" for an attribute, say cost, follow a specific distribution, in this case a normal. In contrast to the MNL and other discrete choice models discussed thus far, the use of a distribution allows the analyst to investigate the hypothesis that some individuals' (facing the same product choices in the market and/or exhibiting similar socio-demographic characteristics) are more price-conscious than other individuals. That is, whereas the MNL and other discrete choice models belonging to the GNL and NetGEV families capture the *average* price sensitivity across the population or clearly defined market segment, the mixed MNL provides information on the *distribution* of individuals' price sensitivities.

From an optimization perspective, the analyst needs to solve for the parameters of a mixed MNL model that define the distribution using numerical approximation. Figure 6.1 approximates the standard deviation associated a normal distribution using four (non-random) draws or support points. The normal distribution shown in the figure has a mean of zero and a standard deviation of three. The vertical lines divide this distribution into five equal parts, which when plotted on a cumulative distribution function represent values or "draws" of {0.2, 0.4, 0.6, 0.8}. The probability of individual *n* choosing alternative *i* would be approximated by averaging four MNL probabilities calculated with these draws: one utility function uses a $\boldsymbol{\beta}$ value associated with cost of -2.52, whereas the other three use $\boldsymbol{\beta}$ values of -0.76, 0.76, and 2.52, respectively.[1]

It is important to note that although this example uses four non-random support points, in application, the analyst needs to consider how many draws to use for each observation, as well as how to generate these draws. However, the process of translating draws (representing cumulative probabilities on the (0,1) interval) into specific $\boldsymbol{\beta}$ values is identical to that presented in the example. The only difference is that instead of draws on the unit interval being pre-determined, random, pseudo-random, or other methods are used. It should also be noted that in application, it is also common for the analyst to investigate different types of parametric distributions (normal, truncated normal, lognormal, uniform, etc.) or non-parametric distributions to see which best fits the data.

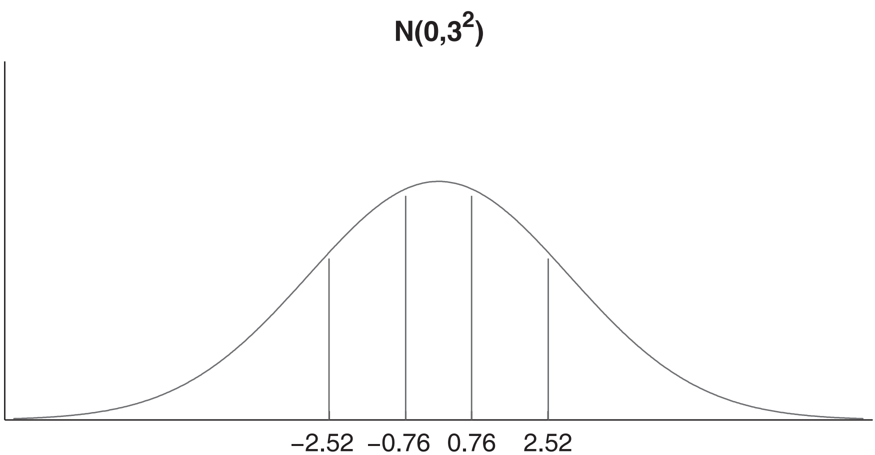


Figure 6.1 Normal distributions with four draws or support points

1 Note that this example assumes the distribution is centered at zero for assigning the "weights" associated with a particular variable (e.g., cost). The center of the distribution, or mean would also be estimated as part of the estimation procedure, but has been suppressed from the example.

Formally, maximum likelihood estimators can be used to solve simultaneously for the fixed $\boldsymbol{\beta}$ coefficients and distribution parameters $\boldsymbol{\eta}$ associated with the random $\boldsymbol{\beta}$ coefficients. Because the integral in Equation (6.1) cannot be evaluated analytically, numerical approximation is used to maximize the simulated maximum likelihood function. The average probability that individual n selects alternative i is calculated by noting that for a particular realization of $\boldsymbol{\beta}$, the logit probability is known. Formally, the average simulated probability is given as:

$$\hat{P}_{ni} = \hat{P}\left(i \mid x_{ni}, \boldsymbol{\beta}, \boldsymbol{\eta}\right) = \frac{1}{R}\sum_{r=1}^{R} L_{ni}\left(\boldsymbol{\beta}_r\right)$$

where:

R is the number of draws or support points used to evaluate the integral,

$\hat{P}_{ni}$ is the average probability that individual n selects alternative i given attributes x_{ni} and parameter estimates $\boldsymbol{\beta}$, which are random realizations of a density function. The parameters of this density function are given by $\boldsymbol{\eta}$,

$\boldsymbol{\beta}_r$ is the vector of parameter estimates associated with draw or support point r.

The corresponding simulated likelihood (SL) and simulated log likelihood (SLL) functions are:

$$SL\left(\boldsymbol{\beta}\right) = \prod_{n=1}^{N} \prod_{i \in C_n} \hat{P}\left(i \mid x_{ni}, \boldsymbol{\beta}, \boldsymbol{\eta}\right)^{d_{ni}}$$

$$SLL\left(\boldsymbol{\beta}\right) = \sum_{n=1}^{N} \sum_{i \in C_n} d_{ni} \ln \hat{P}\left(i \mid x_{ni}, \boldsymbol{\beta}, \boldsymbol{\eta}\right)$$

where:

d_{ni} is an indicator variable equal to 1 if individual n selects alternative i and 0 otherwise.

Mixed GEV Models

At this point in the discussion, before presenting the error components interpretation of the mixed logit model, it is useful to describe an extension of the formulation given in Equation (6.1) and to present an example. The extension involves relaxing the assumption that the core probability embedded in the simulated log likelihood function is a MNL. That is, in the random coefficients interpretation of the mixed logit model, the utility function was defined as

$U_{ni} = \boldsymbol{\beta}' x_{ni} = \varepsilon_{ni}$ and the vector of error components, $\boldsymbol{\varepsilon}$, was implicitly assumed to be IID Gumbel, resulting in a core MNL probability function for $L_{ni}(\boldsymbol{\beta})$. However, as discussed in earlier chapters, different logit models belonging to the GEV class can be derived by relaxing the independence assumption. These same relaxations can be applied in the context of the mixed model, effectively replacing the core MNL probability with a NL, GNL, or other probability function that can be analytically evaluated. That is, just as a NL, GNL, or other GEV model was derived through relaxations of the independence assumption, so too can "mixed NL," "mixed GNL," or "mixed GEV" models be derived.[2] In this manner, the analyst can incorporate random taste variation by allowing the $\boldsymbol{\beta}$ parameters to vary while simultaneously incorporating desired substitution patterns by using different probability functions for $L_{ni}(\boldsymbol{\beta})$.

The advantage of using mixed GEV models to incorporate both random taste variation and correlation among alternatives is clearly seen in the context of the complex two-level and three-level itinerary choice models highlighted in Chapter 5. In this case, it would be undesirable to create dozens—if not hundreds—of mixture error components to approximate these complex substitution patterns when exact probabilities that do not involve numerical approximations (such as those summarized in Table 5.2) can be used.

An example of a mixed NL model based on airline passengers' no show and early standby behavior is shown in Table 6.3. The column labeled "NL" reports the results of a standard nested logit model. The columns labeled "Mix NL 250 Mean" and "Mix NL 500 Mean" reports the results of Mixed NL model that assumes alternative-specific parameters associated with individuals traveling as a group follow a normal distribution; the numbers 250 and 500 indicate whether 250 draws or 500 draws were used. These columns report average parameter estimates obtained from multiple datasets generated from the same underlying distributions. These multiple datasets are typically referred to as *replicates* within the simulation literature. The datasets are identical, except they use different support points for numerical approximation, e.g., for pseudo-random draws, this would be equivalent to using different random seeds to create multiple datasets.

The stability of parameter estimates can be observed by comparing mean parameter estimates and log likelihood functions for those runs based on 250 draws with those runs based on 500 draws. The largest differences in parameter estimates is seen with the group variables, which on average differ by at most 0.003 units for parameters significant at the 0.05 level, and by at most 0.021 units for parameters that are not significant at the 0.05 level. The average log likelihood functions for these two columns are also similar and differ by 0.03 units. The relative stability in parameter estimates can also be observed from the "Mix NL 250 SD" and

2 In the literature, it is common to use the term "mixed model" to refer to a "mixed MNL model," that is, the use of a MNL probability function is implied unless explicitly indicated otherwise.

Table 6.3 Mixed logit examples for airline passenger no show and standby behavior

	NL	Mix NL 250 Mean	Mix NL 250 SD	Mix NL 500 Mean	Mix NL 500 SD	NL Mix 500 Mean	NL Mix 500 SD
Alternative specific constant for NS	1.20 (9.4)	1.302 (7.9)	0.00139	1.302 (7.9)	0.00107	1.789 (9.5)	0.0043
ASC for ESB: Duration ≤ 180 mins	0.17 (1.7)	0.180 (1.5)	0.00015	0.180 (1.5)	0.00011	0.185 (1.1)	0.0004
ASC for ESB: 180 < duration ≤ 300 mins	0.04 (0.3)	0.048 (0.4)	0.00020	0.048 (0.4)	0.00014	0.054 (0.3)	0.0007
ASC for ESB: Duration > 300 mins	-0.38 (3.0)	-0.382 (2.4)	0.00023	-0.382 (2.4)	0.00028	-0.375 (1.8)	0.0004
Alternative specific constant for LSB	-0.20 (2.5)	-0.197 (2.1)	0.00021	-0.197 (2.1)	0.00015	-0.269 (2.1)	0.0007
E-ticket NS	-1.48 (20.)	-1.514 (16.)	0.00100	-1.514 (16.)	0.00072	-2.119 (7.4)	0.0160
Booking Class (ref. = low yield)							
First and business NS	-0.01 (0.1)	-0.052 (0.3)	0.00024	-0.052 (0.3)	0.00019	-0.073 (0.3)	0.0048
First and business ESB	-0.80 (6.9)	-0.817 (4.6)	0.00045	-0.817 (4.6)	0.00045	-1.121 (5.4)	0.0008
First and business LSB	-1.11 (5.8)	-1.143 (4.6)	0.00079	-1.143 (4.6)	0.00068	-1.569 (5.6)	0.0003
High yield NS	0.21 (2.3)	0.103 (0.9)	0.00003	0.103 (0.9)	0.00001	0.139 (0.9)	0.0010
High yield ESB	0.07 (1.1)	0.064 (0.8)	0.00014	0.064 (0.8)	0.00010	0.090 (0.8)	0.0003
High yield LSB	-0.05 (0.7)	-0.062 (0.7)	0.00008	-0.062 (0.7)	0.00007	-0.086 (0.7)	0.0001
Group Size (ref. = travel alone)							
Groups of 2-10 individuals NS mean	-0.35 (4.2)	-0.685 (3.2)	0.00579	-0.686 (3.2)	0.00394	-0.934 (3.2)	0.0068
Groups of 2-10 individuals NS std. dev		0.964 (2.0)	0.01907	0.967 (2.0)	0.02239	1.307 (1.9)	0.0238
Group of 2-10 individuals ESB mean	-0.44 (5.0)	-0.457 (4.1)	0.00718	-0.456 (4.2)	0.00153	-0.629 (4.9)	0.0023
Group of 2-10 individuals ESB std. dev		0.004 (0.1)	0.06759	-0.019 (0.1)	0.02567	0.009 (0.1)	0.1066
Group of 2-10 individuals LSB mean	-0.23 (2.5)	-0.238 (2.4)	0.00079	-0.238 (2.4)	0.00084	-0.328 (2.5)	0.0006
Group of 2-10 individuals LSB std. dev		0.017 (0.1)	0.02353	0.015 (0.1)	0.02111	0.019 (0.1)	0.0369
Logsum	0.71 (3.9)	0.727 (8.3)	0.00045	0.727 (8.2)	0.00037		
NL Mixture						1.109 (3.8)	0.0190
Model Fit Statistics (OBS=3,674 ; LL Zero= -4798; LL Constants = -4681)							
LL Model	-4155	-4150.25	0.067	-4150.22	0.041	-4150.29	0.190
Rho-Squarezero / Rho-Squareconstant	0.134 0.122	0.134 0.113		0.135 0.113		0.135 0.113	

Notes: ASC = alternative specific constant; NS=no show; ESB=early stand-by; LSB=late stand-by; SD or std. dev = standard deviation. Only a subset of parameter estimates are shown; full model results are in Garrow (2004). With exception of NL model, each column is based on approximately 10 runs or separate datasets.

Source: Modified from Garrow 2004: Tables A1.2, A1.6 and A1.7 (reproduced with permission of author).

"Mix NL 500 SD" columns that report standard deviation in parameter estimates obtained from ten different datasets. These columns indicate that the variability in parameter estimates across multiple datasets is small.

To summarize, assessing any changes in parameter estimates and the log likelihood function by using multiple datasets and increasing the number of draws are two strategies analysts can—and should—use to verify that they have used a sufficient number of draws for their particular problem context; failure to use a sufficient number of draws can result in empirical identification problems. Once the stability in parameter estimates has been verified, model parameters can be interpreted. In this example, using random coefficients for the group indicator variables (assumed to follow a normal distribution) suggests that a random distribution may only be helpful in describing no show behavior (versus early or late standby behavior). Note that the means for the group variables associated with the early standby and late standby variables are very similar to those obtained with the NL model, and more importantly, the standard deviation parameter associated with the normal distribution is very small and insignificant. In contrast, the mean parameter estimate associated with the group no show variable is more negative (-0.69 vs. -0.35) in the mixed NL model, and the standard deviation associated with its normal distribution (0.96) is significant at the 0.05 level. Thus, individuals traveling in groups exhibit variability in their no show behavior. Although in general, these individuals are more likely to show than individual business travelers, there is variation in how likely they are to show. This may be due in part to the fact that group size is a proxy for leisure travelers and/or that small group sizes may exhibit different behavior than larger group sizes, which is currently not captured in the utility function.

Error Component Interpretation for Mixed Logit Models

As noted earlier, different interpretations arise for mixed logit models depending on whether $\boldsymbol{\beta}$ varies across individuals, observations, and/or alternatives. When $\boldsymbol{\beta}$ varies across individuals, mixed logits are said to incorporate random taste variation or random coefficients. When $\boldsymbol{\beta}$ varies across observations or alternatives mixed logits are said to incorporate error components. For example, when multiple responses are elicited from the same individual from a survey and/or when the estimation dataset represents panel data, $\boldsymbol{\beta}$ can vary across observations, thereby capturing common unobserved error components or covariance associated with eliciting multiple responses from a single individual. Similarly, when $\boldsymbol{\beta}$ varies across alternatives, mixed logits incorporate error components that enable flexible substitution patterns. These flexible substitution patterns are created by defining x in a manner that creates covariance and/or heteroscedasticity among alternatives. In this manner, analogs to closed-form models can be created via including appropriately defined error components that vary in specific ways across alternatives.

Formally, in an error components derivation, the utility individual n obtains from alternative i is given as $U_{ni} = \boldsymbol{\beta}' x_{ni} + \omega_i(\Xi) + \varepsilon_{ni}$ where $\boldsymbol{\beta}$ is the vector of parameters associated with attributes x_{ni}, ε_{ni} is a random error component, and ω_i is an additional error component (or set of additional error components) associated with alternative i. The additional error components are constructed from an underlying vector of random terms with zero mean given by Ξ.

Although error components are typically used in conjunction with random taste variation, to visualize the equivalence of the error components and random coefficients formulations, assume that $\boldsymbol{\beta}$ is fixed. Similar to the random coefficients formulation, mixed logit choice probabilities are computed as:

$$P_{ni} = \int L_{ni}\left(\boldsymbol{\beta}, \omega_i(\Xi)\right) f(\Xi|\boldsymbol{\eta}) d\Xi \tag{6.2}$$

where:

P_{ni}	is the probability individual n chooses alternative i,
$L_{ni}\left(\boldsymbol{\beta}, \omega_i(\Xi)\right)$	is a logit probability evaluated at the vector of fixed parameter estimates $\boldsymbol{\beta}$ and error component(s), ω_i, that are random realizations from the density function $f(\Xi \mid \boldsymbol{\eta})$,
$\boldsymbol{\eta}$	are parameter estimates of the density function for $f(\Xi)$.

The equivalence of the random coefficients formulation, given in Equation (6.1) with the error components formulation, given in Equation (6.2) is straightforward. Conceptually, the only difference is that in the random coefficients formulation, the coefficients that are randomly distributed are associated with "typical" variables—travel time, travel cost, alternative specific constants, frequent flyer status, etc.—whereas in the error component formulation, the coefficients that are randomly distributed are associated with new "indicator variables" that create specific correlation patterns for sets of alternatives. For example, if two alternatives share a common nest, an indicator variable that is "common" to each of these alternatives is defined. The indicator variable is assumed to follow a standard normal distribution (since the normal closely approximates the Gumbel). The parameter estimate associated with the standard deviation of this indicator variable (i.e., error component), provides a measure of the degree of correlation, or positive covariance, between the two alternatives that share a common nest.

As a more concrete example, consider the NL model in Figure 6.2. Analogs to the NL model can be created via mixed logit error components via $\omega_i(\Xi)$. These analogs are designed to replicate the same pattern of correlation of a pure NL model while using a MNL logit probability for $L_{ni}\left(\boldsymbol{\beta}, \omega_i(\Xi)\right)$. The NL model is approximated using a structure that adds error components to the utility of alternatives that are considered to be part of a common nest to induce correlation among these alternatives (e.g., see Revelt and Train 1998; Brownstone and Train 1999; Munizaga and Alvarez-Daziano 2001, 2002; and Cherchi and Ortuzar 2003). Formally, the added error components in these studies are expressed as:

$$\omega_i\left(\Xi\right)=\sum_{m=1}^{M}\Xi_m d_{mi}=\sum_{m=1}^{M}\sigma_m\cdot\xi_m\cdot d_{mi}$$

where $\omega_i\left(\Xi\right)$ is the additional error component associated with alternative i, d_{mi} is an indicator variable equal to one if alternative i is in nest m and zero otherwise, and Ξ_m are random variables assumed to be iid and follow a normal distribution with mean 0 and variance σ_m^2, iid $N\left(0,\sigma_m^2\right)$. Ξ_m can be rewritten as $\sigma_m \times \zeta_m$ where ζ_m are random variables assumed to be iid and follow a standard normal distribution, and σ_m is a scale parameter that enters the utility of each alternative in nest m. The scale parameter, σ_m, determined during the estimation procedure, represents the standard deviation of the scaled random term and captures the magnitude of correlation among alternatives in nest m. The variance-covariance matrix associated with the NL mixture analog shown in Figure 6.2 is given as:

$$\Omega=\begin{array}{c}1\\[1em]2\\[1em]3\\[1em]4\end{array}\begin{bmatrix}\dfrac{\pi^2}{6\gamma^2} & 0 & 0 & 0\\[1em] & \sigma_1^2+\dfrac{\pi^2}{6\gamma^2} & \sigma_1^2 & \sigma_1^2\\[1em] & & \sigma_1^2+\dfrac{\pi^2}{6\gamma^2} & \sigma_1^2\\[1em] & & & \sigma_1^2+\dfrac{\pi^2}{6\gamma^2}\end{bmatrix}$$

However, it is important to note that this NL error component analog shown in Figure 6.2, commonly used in the literature, introduces both correlation and error heteroscedasticity—that is, the diagonal of the variance covariance is no longer the same across all alternatives and does not maintain the pure NL model assumption that total error for each alternative is identically distributed. This point has been noted by several researchers including Walker (2002), Munizaga and Alvarez-Daziano (2001, 2002), Bhat and Gossen (2004), and Cherchi and Ortuzar (2003). If desired, additional error components can be used to allow for correlation only and maintain equal variance across alternatives; e.g., see Garrow and Bodea (2005) and Bodea and Garrow (2006) for examples.

A numerical example based on the NL model structure shown in Figure 6.2 is contained in Table 6.3. In this example, an error component is created for the show, early standby, and late standby alternatives that share a common nest. This is accomplished through defining an indicator variable equal to one if the alternative is show, early standby or late standby. The indicator variable is assumed to follow a normal distribution with mean zero and standard deviation

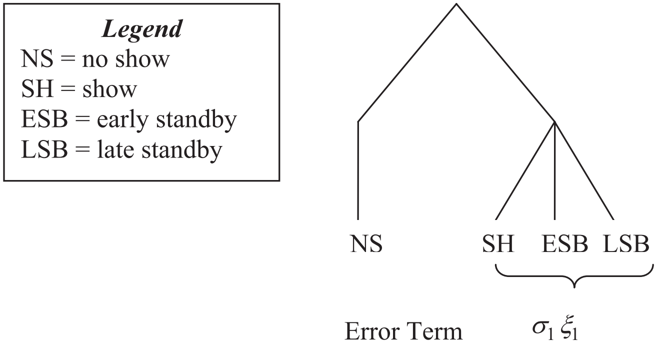


Figure 6.2 Mixed error component analog for NL model

that is estimated from the data. As seen in Table 6.3 for the "NL Mix 500 Mean" column, this NL mixture model results in an estimated standard deviation of 1.109. A comparison of the NL mixture model with the Mixed NL model reveals that whereas the mean log likelihood values are similar (-4150.29 versus -4150.22) the variability in parameter estimates for the NL mixture models is, in general, higher.

To date, there have been several empirical papers that have compared GEV models with one or more mixed logit error component models. For example, Gopinath, Schofield, Walker, and Ben-Akiva (2005), Hess, Bierlaire, and Polak (2005a), and Munizaga and Alvarez-Daziano (2001, 2002) compared GEV and mixed MNL models that included heteroscedastic error components, whereas Munizaga and Alvarez-Daziano (2001, 2002) compared GEV and mixed MNL models that included homoscedastic error components. In general, although theoretically a homoscedastic error component structure more closely approximates a pure NL model (due to maintaining the assumption of equal variance across alternatives), in empirical applications it is more common to use the heteroscedastic error representation.

Estimation Considerations

Although some authors have begun to investigate alternative methods for solving for the parameters of the mixed logit model (e.g., see Guevara, Cherchi and Moreno 2009), there are two key estimation considerations that researchers always need to consider when using the approach outlined in this chapter. These include determining the distribution(s) associated with random coefficients and determining the number and types of draws to be used as support points for numerically evaluating integrals.

Common Mixture Distributions

As shown in Table 6.1, in most of the early mixed logit applications, the density function was assumed to have a normal, truncated normal, or lognormal distribution. Uniform, triangular, and other distributions have also been explored, particularly in the context of modeling individuals' value of time (often represented as the ratio of coefficients associated with time and cost, e.g., $\beta_{time} / \beta_{cost}$). The use of normal distributions for both time and cost variables is undesirable, due to the fact that the distribution of the ratio of two normal random variables is distributed Cauchy, a distribution that may not have a finite mean (e.g., see Hoel Port and Stone 1971). Subsequently, the lognormal has been used, as it has the advantage over the normal distribution (and probit model) in that it ensures a coefficient maintains the same sign across the entire population. This is particularly advantageous in the context of modeling individuals' value of time as the coefficient associated with price is typically assumed to always be negative; that is, the utility associated with an alternative is expected to decrease as the price increases. Alternatively, the truncated normal has the advantage over the normal and lognormal distributions in that it prevents extreme, unrealistic realizations of the utility function associated with the tails of the normal and lognormal distributions. In the context of bounded distributions, Hensher (2006) proposed the use of a global constraint on the marginal disutility, which effectively ensures that the value of time maintains positive values when used with a broad range of distributions (e.g., Hensher provides an empirical example using a globally constrained Rayleigh distribution). In a similar spirit, Train and Sonnier (2005) create bounded distributions of correlated partworths via transformations of joint normal distributions (providing examples using lognormal, censored normal, and Johnson's S_B distribution).

All of these distributions are parametric forms that must be determined a priori by the researcher. Recent papers investigating non-parametric methods for mixed logits include those by Dong and Koppelman (2003), Hess, Bierlaire, and Polak (2005b), Fosgerau (2006), Fosgerau and Bierlaire (2007), Bastin, Cirillo, and Toint (2009), Cherchi, Cirillo, and Polak (2009), and Swait (2009). Nonparametric distributions can be superior to parametric distributions and are particularly helpful in uncovering distributional forms that are unexpected a priori.

To summarize, although most current applications of mixed logits assume normal or lognormal distributions, it is important to recognize that a wide range of parametric and non-parametric distributions can be used. As with the value of time example, the most appropriate distribution will be application-specific and/or data-specific.

Number of Draws for Numerical Approximation

Much of the research in the late 1990's and early 2000's was focused on comparing different quasi-random or low-discrepancy number sequences and determining "how many" and "what type" of draws should be used to approximate multi-

dimensional integrals. As noted by Ben-Akiva, Bolduc, and Walker (2001) in addition to other researchers, the number of draws (or points) necessary to simulate the probabilities with good precision depends on the type of draws, model specification, and data. Indeed, as shown in Table 6.1, upon synthesizing results from multiple applications of mixed logit models, it can be easily observed that there are no unifying guidelines for deciding "how many draws" are enough and how "precision" should be defined. For example, Hensher (2001b) reports stability in model parameters for as few as 50 Halton draws whereas Beckor, Ben-Akiva, and Ramming (2002) report model instability even after using 100,000 draws. Also, whereas Hensher (2001b) measured stability in terms of the ratio of two parameters (i.e., value of time ratios), Beckor, Ben-Akiva, and Ramming (2002) measured stability in terms of individual parameters and overall log likelihood values.

From a research perspective, it is important to highlight two results from the optimization literature related to Monte-Carlo methods used to evaluate multi-dimensional integrals. To date, the importance of these results has not been fully recognized in the transportation literature. Specifically, the optimization literature reveals that the number of draws or support points required to maintain a specified relative error criteria in the objective function, i.e., log likelihood value, *increases exponentially as the dimensionality of the integral being approximated increases* whereas the number of support points required to maintain a specified absolute error criteria in the objective function *increases linearly with increases in the dimensionality of the integral* (e.g., see Fishman 1996: p. 55). Although in practice, using draws that seek to improve coverage can help reduce these upper bounds on error, empirical evidence suggests that deciding "how many" draws are enough is application specific; researchers would be wise not to decide *a priori* the "right" number of draws to use based on prior applications. As noted by Walker (2002), the number of draws must be sufficiently large so that parameter results are stable or robust as the number of draws increases. This, of course, can only be assessed by testing the sensitivity of results to the number of draws.

One of the most convincing arguments on the need to assess the number of draws used in simulation is seen in work by Chiou and Walker (2007), who conduct a study using actual and synthetic datasets that contained either theoretical and/or empirical identification problems. However, when a low number of draws was used, it was possible to obtain parameter estimates that appeared to be identified (when in reality they were not). It is significant to note that the "false" identification results they report occurred with 1000 pseudo-random draws, which as seen in Table 6.1, is much higher than the number of draws typically reported in the transportation literature. Unfortunately, today many studies using mixed logit models do not report the number (or type) of draws used in the study. For example, out of a dozen mixed logit studies presented at a recent meeting of the Transportation Research Board, only three mentioned the number and type of draws that were used (Habib and Miller 2009; Kim, Ulfarsson, Shankar and Mannering 2009; Shiau, Michalek and Hendrickson 2009); an additional two

studies mentioned only the type of draws used (Eluru, Senar, Bhat, Pendyala and Axhausen 2009; Sener, Eluru and Bhat 2009).

Types of Draws for Numerical Approximation

Some of the earliest research involving mixed logit applications was focused on finding more efficient ways to solve for parameters. Early mixed logit applications using simulation techniques approximated the integral in Equation (6.1) using *pseudo-random* draws. The term *pseudo-random* is used to highlight the distinction between draws generated from a "purely random" process (such as the roll of a die or flip of a coin) and draws that are generated from a mathematical algorithm. The mathematical algorithm is designed to mimic the properties of a pure random sequence (but also provides an advantage in that multiple researchers can generate "identical" random draws).

Currently, most mixed logit applications evaluate the integral in Equation (6.1) using *variance-reduction techniques.* These techniques generate draws from the mixing distribution in a manner that seeks to improve coverage and induce negative correlation over observations, thereby "reducing variance" in the simulated log likelihood function. As an example, compare the three panels in Figure 6.3. The two upper panels each contain 500 (x,y) pairs that were pseudo-randomly generated. When random or pseudo-random draws are used, it is common to have certain areas that contain more pairs (or exhibit greater *coverage*) than other areas, which subsequently increases the variance associated with the simulated log likelihood function. In the upper panel, the right circle contains 14 points whereas the left circle (of equal area) contains no points. In the middle random panel, the left circle contains 22 points whereas the right circle contains three points. Variance-reduction techniques, such as those based on Halton sequences shown in the bottom panel of Figure 6.3, can be used to help distribute points more "evenly" throughout the space, thereby avoiding poor coverage in certain areas and high coverage in others. The three circles in the bottom panel contain between six and nine pairs.

One of the most popular methods for generating pseudo-random draws is based on a method developed by Halton in 1960. The popularity of the Halton method applies not only for mixed logit applications, but to a broad range of simulation applications. A Halton sequence is generated from a prime number. For example, given a utility function with three random coefficients to be estimated, an analyst would create three separate Halton sequences, one associated with each random coefficient, using three prime numbers (e.g., two, three, and five). Figure 6.4 illustrates how Halton draws are generated on a unit interval using the prime number of two.

The generation of Halton draws can be visualized in Figure 6.4 by reading the chart from top to bottom, and using the line definitions provided in the legend to visualize how draws are generated within a given row. In the first "row," indicated by $2^1 = 2$, a single Halton draw is generated at the point 1/2. It is useful to visualize this first "row" as dividing the unit interval into two distinct segments (represented by the vertical line emanating from the point 1/2). The first segment (or "left panel")

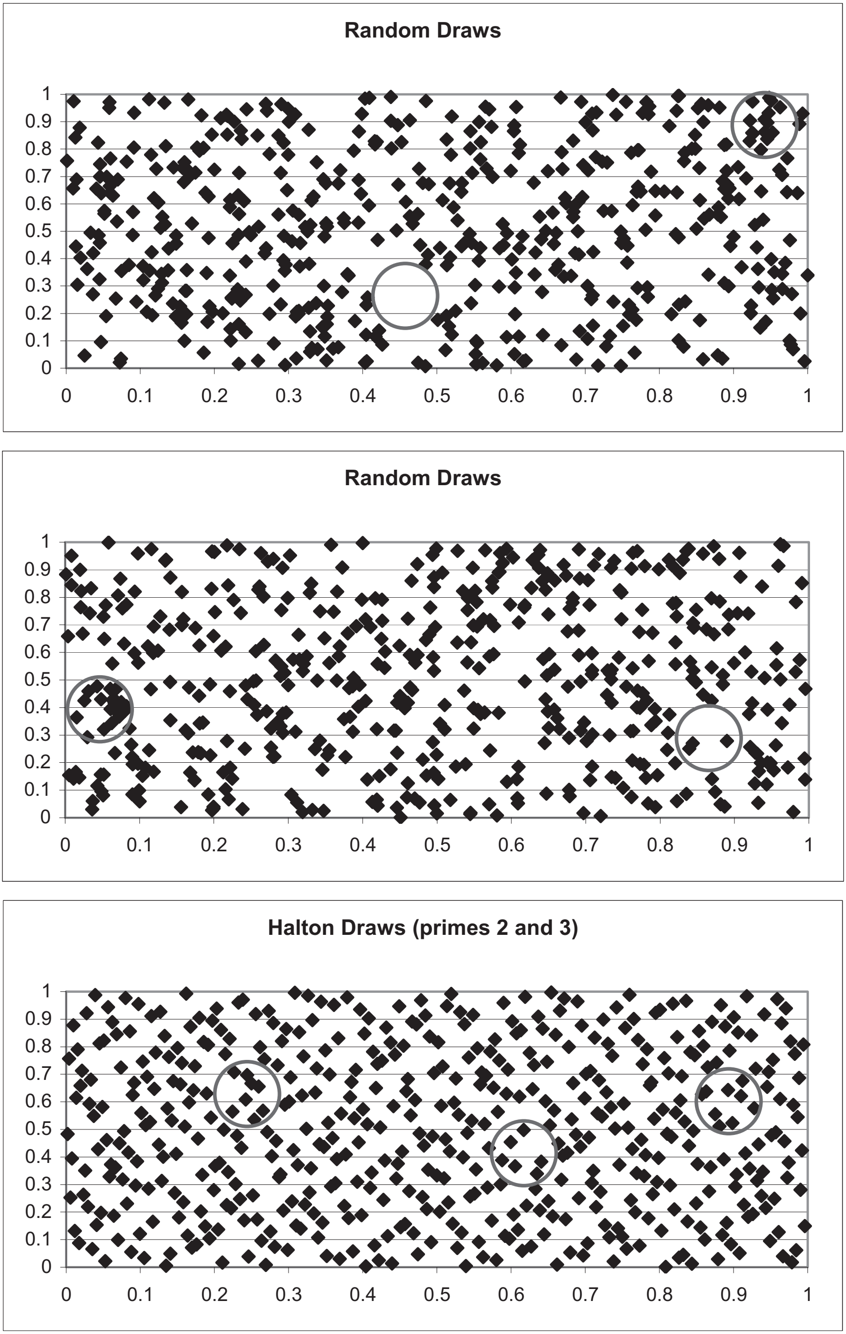


Figure 6.3 Comparison of pseudo-random and Halton draws

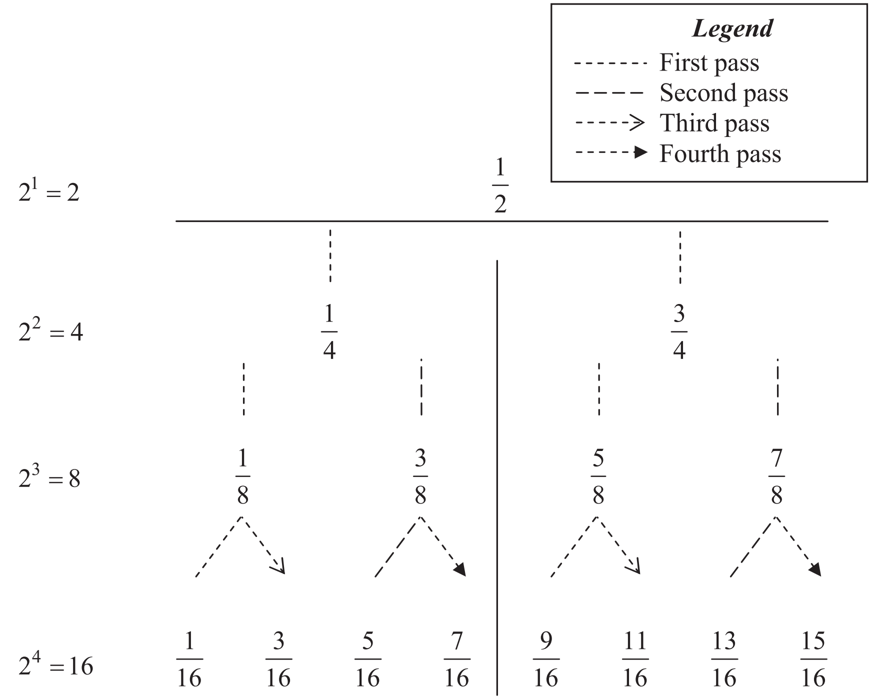


Figure 6.4 Generation of Halton draws using prime number two

represents those points contained in the unit interval that are less than 0.5 and the second segment (or "right panel") represents those points contained in the unit interval that are greater than 0.5. The second "row," indicated by $2^2 = 4$, effectively divides these two segments into four segments, i.e., by first generating a point on the left panel at 1/4 and then generating a point on the right panel at 3/4. Similarly, the third "row," indicated by $2^3 = 8$, effectively divides these four segments into eight segments. Note that in populating the points for the third row, there are two "passes." For the first pass, the points corresponding to the short dashed line are populated, 1/8 and 5/8, whereas for the second pass, points corresponding to the long dashed line, 3/8 and 7/8, are populated. Finally, the fourth "row," indicated by $2^4 = 16$, further divides the eight segments into 16. Note that identical to the third row, two points (corresponding to prime number two) are populated per pass. One of these points is always on the left panel, while the other is always on the right panel. Within a panel, the points to the left of the previous row are first populated, followed by the points to the right of the previous row. This relationship is portrayed using the "tree." For the first pass, the left points of the tree 1/16 and 9/16 are populated. For the second pass, more left tree points remain, and thus 5/16 and 13/16 are populated. For the third pass, "right" points 3/16 and 11/16 are populated, followed by "right" points 7/16 and 15/16. The process repeats until the desired number of draws (or support points)

is obtained. To summarize, the Halton draws for prime number two are generated according to the following order:

$$\begin{array}{llllllll}
\frac{1}{2} \\
\frac{1}{4} & \frac{3}{4} \\
\frac{1}{8} & \frac{5}{8} & \frac{3}{8} & \frac{7}{8} \\
\frac{1}{16} & \frac{9}{16} & \frac{5}{16} & \frac{13}{16} & \frac{3}{16} & \frac{11}{16} & \frac{7}{16} & \frac{15}{16}
\end{array}$$

The generation of Halton draws is very similar for other prime numbers. Figure 6.5 extends the data generation process outlined above for the prime number three. In the case, the key conceptual difference is that instead of originally dividing the

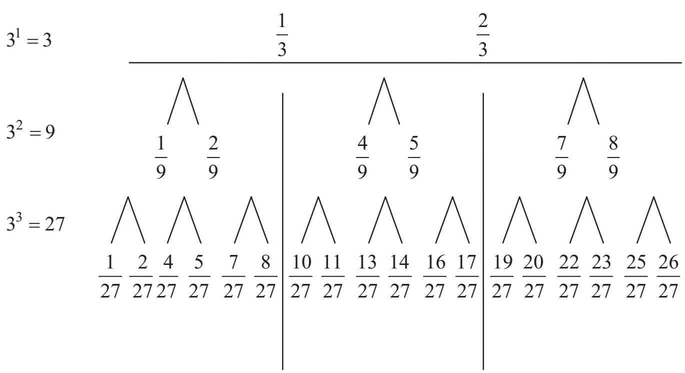


Figure 6.5 Generation of Halton draws using prime number three

unit interval into two panels, three panels are now created (given by the points 1/3 and 2/3) and for each pass, three points (versus two) are populated. Thus, for the second "row," indicated by $3^2 = 9$, the three segments are divided into nine, by first generating the "left pass" points for the left panel at 1/9 the middle panel at 4/9 and the right panel at 7/9. This is followed by the "right pass" points at 2/9, 5/9 and 8/9. The process is repeated for the third row, with the order of Halton draws associated with prime three for the first three rows given as:

$$\begin{array}{llllllllllllllllll}
\frac{1}{3} & \frac{2}{3} \\
\frac{1}{9} & \frac{4}{9} & \frac{7}{9} & \frac{2}{9} & \frac{5}{9} & \frac{8}{9} \\
\frac{1}{27} & \frac{10}{27} & \frac{19}{27} & \frac{4}{27} & \frac{13}{27} & \frac{22}{27} & \frac{7}{27} & \frac{16}{27} & \frac{25}{27} & \frac{2}{27} & \frac{11}{27} & \frac{20}{27} & \frac{5}{27} & \frac{14}{27} & \frac{23}{27} & \frac{8}{27} & \frac{17}{27} & \frac{26}{27}
\end{array}$$

A final example is shown in Figure 6.6 for the generation of Halton draws using prime number five. In this case, the unit interval is originally divided into five "panels" and the tree emanating from each panel contains four branches (effectively used to subdivide each sub-interval created on the previous row into five new sub-intervals, e.g., create 25 intervals as part of row 2 given by 5^2, 125 intervals as part of row 3 given by 5^3, etc. Similar constructs and processes are used for each prime number.

Although the process for generating Halton draws is straightforward, problems can arise when using Halton draws to evaluate high-dimension integrals. Conceptually, this is because Halton draws generated with large prime numbers can be highly correlated with each other. This problem is depicted in Figure 6.7, which contains 500 draws associated with prime numbers 53 and 59, which correspond to the 16th and 17th prime numbers, respectively. Figure 6.7 also illustrates another subtle issue that arises when the number of draws selected (in this case 500) is not a multiple of the prime number used to generate the draws. In this case, poor coverage is exhibited, as seen by the fact that one of the lines "unexpectedly ends." Conceptually, this occurs because draws from the last "row" have not been fully populated, i.e., in the case of prime number 53, the second row is used to generate $53^2 = 2809$ points, but the figure shows only the first 500 points.

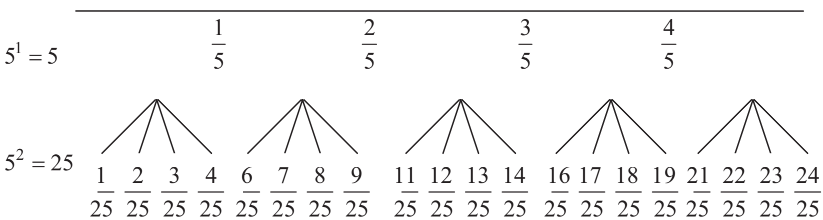


Figure 6.6 Generation of Halton draws using prime number five

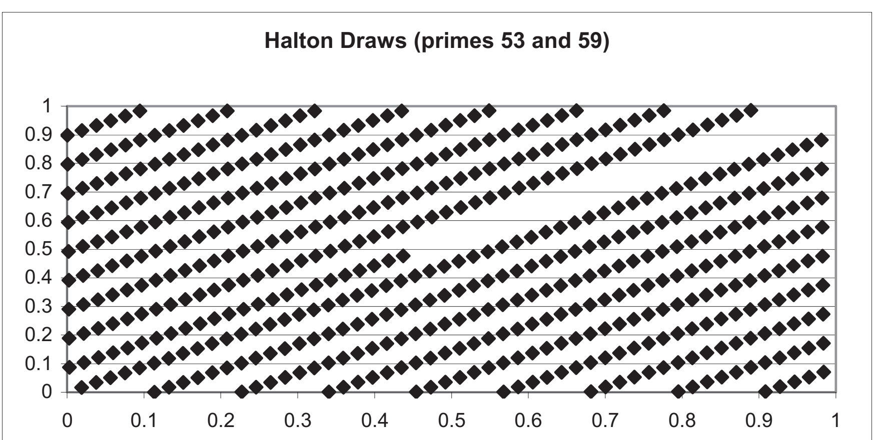


Figure 6.7 Correlation in Halton draws for large prime numbers

Multiple techniques are available to help decrease correlation among draws used to evaluate high-dimension integrals. These techniques include scrambling (which effectively changes the order of how draws are populated within a given row) and randomization techniques. Randomization techniques can loosely be thought of as generating points based on a systematic process, such as Halton sequences, and then adding noise to each point so that the desired coverage structure is maintained, but points are randomly shifted (usually close to where they were generated) to help decrease the high correlation.

Indeed, just as it is important to explore how many draws should be used for a specific application, it is important to decide "what type" of draws should be used. Further, answering "what type" of draws is a question some researchers spend their entire careers investigating. Numerous quasi-Monte Carlo methods have been developed, some of which have been explored in the mixed logit context. Lemieux, Cieslak, and Luttmer (2004) provide an excellent overview of many of these methods, which they have implemented in the C programming language; their code is freely available online. The quasi-Monte Carlo methods they have implemented include the following: Halton sequences, randomized generalized Halton sequences, Sobol's sequence, generalized Faure sequences, the Korobov method, Polynomial Korobov rules, the Shift-net method, Salzburg Tables, Modified Latin Hypercube Sampling, and Generic Digital Nets. The code also includes several randomization techniques, including adding a shift of modulo 1, addition of a digital shift in base b, and randomized linear scrambling.

Within the mixed logit modeling context, researchers have compared the performance of pseudo-random draws and draws based on Halton sequences (Halton 1960), Sobol sequences, and (t,m,s)-nets. Early work in this area includes that of Bhat (2001), Hensher (2001b), and Train (2000) who examined Halton sequences. Because the standard Halton exhibits poor coverage in high dimensions of integration (which in the discrete choice literature may be loosely thought of in terms of 15 or more dimensions), research has expanded beyond using standard Halton sequences to identify methods that help improve coverage of higher-integral domains. In the context of econometric models, randomized and scrambled Halton sequences (based on scrambling logic proposed by Braaten and Weller, 1979) have been examined by Bhat (2003b), Halton sequences based on randomly shifted and shuffled Halton sequences have been examined by Hess, Train and Polak (2004), (t,m,s)-nets have been examined by Sandor and Train (2004), Sobol sequences have been examined by Garrido (2003), and Modified Latin Hypercube Sampling has been examined by Hess, Train, and Polak (2006). Although a detailed comparison of results is not provided here, it is nonetheless interesting to note that similar to the empirical results observed in the context of "how many" draws should be used, no consistent picture of "which draws" should be used has emerged from the literature. Currently, however, most applications of mixed logit models within transportation are based on the pure Halton draws. For example, out of the five papers presented at the 2009 meeting of the Transportation Research Board that mentioned the types of draws used, one used random draws (Shiau, Michalek and Hendrickson 2009), one

used Scrambled Halton draws (Eluru, Senar, Bhat, Pendyala and Axhausen 2009), and three used Halton draws (Habib and Miller 2009; Sener, Eluru and Bhat 2009; Kim, Ulfarsson Shankar and Mannering 2009).

Most important, the theoretical and practical implications of using low-discrepancy sequences versus pure Monte Carlo techniques to date have not been explicitly acknowledged in the transportation literature, aside from a few exceptions like the discussion by Bastin, Cirillo, and Toint (2003). Their observations, which highlight several key underlying theoretical issues, are summarized below. With regard to the recent trend of using low-discrepancy sequences, Bastin, Cirillo, and Toint (2003) observe that:

> "The trend is not without drawbacks. For instance, Bhat (2001) recently pointed out that the coverage of the integration domain by Halton sequences rapidly deteriorates for high integration dimensions and consequently proposed a heuristic based on the use of scrambled Halton sequences. He also randomized these sequences in order to allow the computation of the simulation variance of the model parameters. By contrast, the dimensionality problem is irrelevant in pure Monte-Carlo methods, which also benefit from a credible theory for the convergence of the calibration process, as well as of stronger statistical foundations … In particular, statistical inference on the optimal value is possible, while the quality of results can only be estimated in practice, (for procedures based on low-discrepancy sequences), by repeating the calibration process on randomized samples and by varying the number of random draws."

To summarize, the main advantage of using low-discrepancy sequences is that fewer draws per simulation are generally required. However, this advantage may be outweighed by two key considerations. First, the use of low-discrepancy sequences may not be appropriate for high dimensions of integration due to their inherent poor coverage. Second, unlike pure Monte-Carlo methods, statistical inference on the optimal log likelihood value (e.g., bias and accuracy measures) is not possible; stated another way, the researcher may need to conduct more overall simulation runs using low-discrepancy sequences to obtain accurate numerical approximations of simulation error. Thus, although current research has been centered on applying low-discrepancy sequences, another area of research is to develop more efficient optimization approaches that use pseudo-random sequences.

Identification

In earlier chapters, proper identification and normalization of discrete choice models appeared in several contexts. For example, in Chapter 2, the fact that only differences in utility are uniquely identified was shown to impact how variables that do not vary across the choice set need to be included in the utility function (e.g., when including alternative-specific constants, it is common to normalize the model by setting

one constant to zero). This was also shown to lead to the need for normalization requirements on error assumptions (e.g., it is common to set the scale parameter of the Gumbel to one). As the discussion of model structures became more complex, so too did the underlying normalization rules, as seen in the discussion of the "crash-free" and "crash-safe" rules developed for the NetGEV model in Chapter 5.

The development of identification and normalization rules for mixed logit models has focused on heteroscedastic error component formulations that seek to incorporate correlation structures among alternatives that are similar to those for NL, GNL, and other two-level models[3] (Walker 2001, 2002; Ben-Akiva, Bolduc and Walker 2001; Walker, Ben-Akiva and Bolduc 2007). Because the application of the rules Walker and her colleagues developed are quite involved, the primary objectives of this section are to provide an overview of these rules and summarize open research questions related to the identification of mixed logit models.

Conceptually, the identification and normalization rules proposed by Walker and her colleagues consist of two main steps. First, the number of identifiable covariance terms is determined using order and rank conditions, which are similar in spirit to those proposed by Bunch (1991) in the context of probit models. Second, verification that a particular normalization is valid is determined using the positive definiteness condition, which is designed to ensure that a particular normalization selected by the analyst does not result in negative covariance terms.

The application of the first step is straight-forward and can be visualized via an example. Figure 6.8 portrays a NL model that has five alternatives and two nests. Defining γ is the scale parameter associated with the Gumbel distribution and z as $\pi^2/6$, the variance-covariance matrix associated with this model is given as:

$$\Omega = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{bmatrix} \sigma_1^2 + \dfrac{z}{\gamma^2} & \sigma_1^2 & \sigma_1^2 & 0 & 0 \\ & \sigma_1^2 + \dfrac{z}{\gamma^2} & \sigma_1^2 & 0 & 0 \\ & & \sigma_1^2 + \dfrac{z}{\gamma^2} & 0 & 0 \\ & & & \sigma_2^2 + \dfrac{z}{\gamma^2} & \sigma_2^2 \\ & & & & \sigma_2^2 + \dfrac{z}{\gamma^2} \end{bmatrix}$$

3 These authors have also investigated identification and normalization rules for models that incorporate alternative-specific error components (or include an error component that follows a normal distribution for each alternative). In this case, the authors find that the alternative that has the minimum alternative-specific variance is the one that should be normalized to zero.

Using the identification and normalization rules developed by Walker and her colleagues, it can be shown that this model is not uniquely identified. The order condition is first used to identify the maximum number of alternative-specific error parameters that can be identified using the order condition. The order condition states that for J alternatives, at most $(J \times (J - 1)/2) - 1$ alternative-specific error parameters can be identified. Thus, in the five-alternative example shown in Figure 6.8, at most nine parameters (eight σ covariance components and one variance scale γ) can be identified.

Whereas the order condition provides an upper bound on the number of parameters that can be estimated, the rank condition is more restrictive. The rank condition is based on the covariance matrix of differences in utilities. Using the relationship that:

Cov(A-B, C-B) = Var(B) + Cov(A,C) – Cov(A,B) – Cov(C,B)

the covariance matrix of utility differences relative to alternative five for this example is given as:

$$\Delta\Omega = \begin{array}{c} 1-5 \\ 2-5 \\ 3-5 \\ 4-5 \end{array} \begin{bmatrix} \sigma_1^2 + \sigma_2^2 + \frac{2z}{\gamma^2} & \sigma_1^2 + \sigma_2^2 + \frac{z}{\gamma^2} & \sigma_1^2 + \sigma_2^2 + \frac{z}{\gamma^2} & \frac{z}{\gamma^2} \\ & \sigma_1^2 + \sigma_2^2 + \frac{2z}{\gamma^2} & \sigma_1^2 + \sigma_2^2 + \frac{z}{\gamma^2} & \frac{z}{\gamma^2} \\ & & \sigma_1^2 + \sigma_2^2 + \frac{2z}{\gamma^2} & \frac{z}{\gamma^2} \\ & & & \frac{2z}{\gamma^2} \end{bmatrix}$$

The unique elements in $\Delta\Omega$ can be expressed in vector form as:

$$\begin{bmatrix} \sigma_1^2 + \sigma_2^2 + \frac{2z}{\gamma^2} \\ \sigma_1^2 + \sigma_2^2 + \frac{z}{\gamma^2} \\ \frac{z}{\gamma^2} \\ \frac{2z}{\gamma^2} \end{bmatrix}$$

The Jacobian of this vector with respect to each of the unknown parameters is given as:

$$\begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$

which has a rank of two, which implies that one parameter can be estimated and two parameters (the variance scale γ and one σ) must be constrained.

The second step in the identification and normalization rules developed by Walker and her colleagues is designed to ensure that the normalization selected by the analyst does not change the original variance-covariance matrix (e.g., covariance terms can only be positive by model definition). This is accomplished via the positive definiteness condition, which checks, among other things, that the normalization selected by the analyst maintains non-negative (positive and zero) covariance terms; for a numerical example, see Ben-Akiva, Bolduc, and Walker (2001). These identification and normalization rules, while satisfying both necessary and sufficient conditions, can be tedious to apply. Consequently, Bowman (2004) derived a set of identification and normalization rules for heteroscedastic error component models that are easier to apply and satisfy the necessary (but not sufficient) condition.

In the case of a NL model with two nests, normalization of the covariance terms is arbitrary and can be accomplished by constraining $\sigma_1^2 = \sigma_2^2$ or by setting either σ_1^2 or σ_2^2 to zero. Further, although the heteroscedastic NL analog containing two nests, such as that shown in Figure 6.8, is not uniquely identified, heteroscedastic NL analogs containing one nest or three or more nests are uniquely identified. This result is reported in Walker (2001, 2002), Ben-Akiva, Bolduc, and Walker (2001), and Walker, Ben-Akiva, and Bolduc (2007).

From a research perspective, the development of identification and normalization rules for random coefficients has been less studied. On one hand, it is easy to verify that theoretically, a random coefficient associated with a generic variable (such as travel time or cost) that varies across the choice set and estimation sample is uniquely identified. However, as noted by Ben-Akiva, Bolduc and Walker (2001, p. 28), the issue of identification for "the case when random parameters are specified for multiple categorical variables in the model … is not addressed in the literature" and is an open area of research. That is, although the discrete choice modeling community has clearly embraced mixed logit models and has applied them in numerous decision-making contexts, it is important to note that there are still several fundamental research questions related to identification that remain to be investigated. This includes extension of heteroscedastic error component models for analogs of NetGEV models that contain three or more levels, as well

as extensions to random coefficient models that contain multiple categorical variables.

Summary of Main Concepts

This chapter presented an overview of the mixed logit model. The most important concepts covered in this chapter include the following:

- The mixed logit model is able to relax several assumptions inherent in the GNL and NetGEV models, i.e., it is able to incorporate random taste variation, correlation across observations (in addition to correlation across alternatives), and heteroscedasticity.
- The mixed logit model has been shown to theoretically approximate any random utility model.
- Two common formulations for the mixed logit model include the random coefficients formulation and the error components formulation.
- Conceptually, the mixed logit model is similar to the probit model in that choice probabilities must be numerically evaluated. However, this computation is facilitated by embedding the MNL (or other closed-form GEV model) as the core within the likelihood function.
- The phrase "mixed logit model" is commonly used to refer to a random coefficients logit model that uses a MNL probability to calculate choice probabilities. A "mixed GEV" model replaces the MNL probability with another choice model (NL, GNL, etc.) that belongs to the family of GEV models.
- Although the mixed logit has been embraced by the discrete choice modeling community and has been applied in numerous transportation contexts, applications of the mixed logit model in aviation has been limited to studies based on stated preference data and/or publicly-available data versus proprietary industry datasets.
- Analysts should always make sure to test the stability of model estimation results to the number of support points (or draws) used for numerical approximation.
- Halton draws are commonly used to generate support points for mixed logit models. However, it should be noted that when estimating high-dimensional mixed logit models, alternative variance-reduction techniques need to be investigated because Halton draws generated with large prime numbers can be highly correlated with each other.
- Given that the investigation of the theoretical and empirical identification properties associated with mixed logit models is still an open area of research, it is highly recommended that analysts clearly document simulation details (e.g., number and types of draws used) in publications.

Chapter 7

MNL, NL, and OGEV Models of Itinerary Choice

Laurie A. Garrow, Gregory M. Coldren, and Frank S. Koppelman

Introduction

Network-planning models (also called network-simulation or schedule profitability forecasting models) are used to forecast the profitability of airline schedules. These models support many important long- and intermediate-term decisions. For example, they aid airlines in performing merger and acquisition scenarios, route schedule analysis, code-share scenarios, minimum connection time studies, price-elasticity studies, hub location and hub buildup studies, and equipment purchasing decisions. Conceptually, "network-planning models" refer to a collection of models that are used to determine how many passengers want to fly, which itineraries (defined as a flight or sequence of flights) they choose, and the revenue and cost implications of transporting passengers on their chosen flights.

Although various air carriers, aviation consulting firms, and aircraft manufacturers own proprietary network-planning models, very few published studies exist describing them. Further, because the majority of academic researchers did not have access to the detailed ticketing and itinerary data used by airlines, the majority of published models are based on stated preference surveys and/or a high level of geographic aggregation. These studies provide limited insights into the range of scheduling decisions that network-planning models must support. Recent work by Coldren and Koppelman provide some of the first details into network-planning models used in practice (Coldren 2005; Coldren and Koppelman 2005a, 2005b; Coldren, Koppelman, Kasturirangan and Mukherjee 2003; Koppelman, Coldren and Parker 2008). This chapter draws heavily from the work of Coldren and Koppelman and from information obtained via interviews with industry experts.

This chapter has two primary objectives. The first objective is to provide an overview of the major components of network-planning models and contrast two major types of market share models—one based on the Quality of Service Index (QSI) methodology and the second based on logit methodologies. The second objective is to illustrate the modeling process that is used to develop a well-specified utility function and relax restrictive substitution patterns associated with the MNL model. Based on these objectives, this chapter is organized into several sections. First, an overview of the major components of network-planning models

is presented. This is followed by an in-depth examination of the logit modeling process. Specifically, major statistical tests used to compare different models are first described, followed by development of the MNL, NL, and OGEV itinerary choice models.

Overview of Major Components of Network-Planning Models

As shown in Figure 7.1, "network-planning models" refer to a collection of sub-models. First, an *itinerary generation algorithm* is used to build itineraries between each airport pair using leg-based air carrier schedule data obtained from a source such as the Official Airline Guide (OAG Worldwide Limited 2008). OAG data contain information for each flight including the operating airline, marketing airline (if a code-share leg), origin, destination, flight number, departure and arrival times, equipment, days of operation, leg mileage and flight time. Itineraries, defined as a flight or sequence of flights used to travel between the airport pair, are constructed from the OAG schedule. Itineraries are usually limited to those with a level of service that is either a non-stop, direct (a connecting itinerary not involving an airplane change), single-connect (a connecting itinerary with an airplane change)

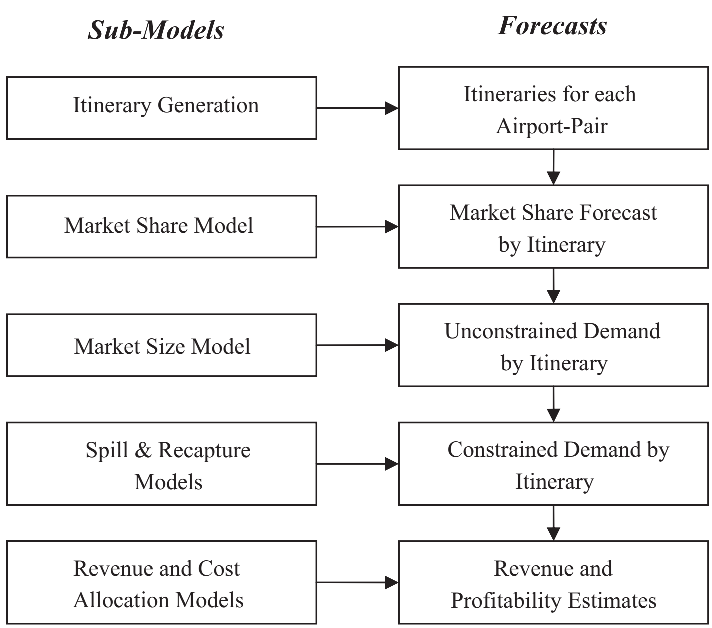


Figure 7.1 Model components and associated forecasts of a network-planning model

or double-connect (an itinerary with two connections). For a given day, an airport pair may be served by hundreds of itineraries, each of which offers passengers a potential way to travel between the airports. Although the logic used to build itineraries differs across airlines, in general itinerary generation algorithms include several common characteristics. These include distance-based circuitry logic to eliminate unreasonable itineraries and minimum and maximum connection times to ensure that unrealistic connections are not allowed. In addition, itineraries are typically generated for each day of the week to account for day-of-week differences in service offered.

An exception to the itinerary generation algorithm described above was developed by Boeing Commercial Airplanes for large-scale applications used to allocate weekly demand on a world-wide airline network. In this application, a weekly airline schedule involves the generation of 4.8 million paths across 280,000 markets that are served by approximately 950 airlines with 800,000 flights. Boeing's algorithms, outlined in Parker, Lonsdale, Glans, and Zhang (2005), integrate discrete choice theory into both the itinerary generation and itinerary selection. That is, the utility value of paths is explicitly considered as the paths are being generated; those paths with utility values "substantially lower" than the best path in a market are excluded from consideration.

After the set of itineraries connecting an airport pair is generated, a *market share model* is used to predict the percentage of travelers that select each itinerary in an airport pair. Different types of market share models are used in practice and can be generally characterized based on whether the underlying methodology uses a QSI or discrete choice (or logit-based) framework. Both types of market share models are discussed in this chapter.

Next, demand on each itinerary is determined by multiplying the percentage of travelers expected to travel on each itinerary by the forecasted *market size*, or the number of passengers traveling between an airport pair. However, because the demand for certain flights may exceed the available capacity, *spill and recapture models* are used to reallocate passengers from full flights to flights that have not exceeded capacity. Finally, *revenue and cost allocation models* are used to determine the profitability of an entire schedule (or a specific flight).

Market size and market share information can be obtained from ticketing data that provide information on the number of tickets sold across multiple carriers. In the U.S., ticketing data are collected as part of the U.S. Department of Transportation (US DOT) O*rigin and Destination Data Bank 1A or Data Bank 1B* (commonly referred to as DB1A or DB1B). The data are based on a 10 percent sample of flown tickets collected from passengers as they board aircraft operated by U.S. airlines. The data provide demand information on the number of passengers transported between origin-destination pairs, itinerary information (marketing carrier, operating carrier, class of service, etc.), and price information (quarterly fare charged by each airline for an origin-destination pair that is averaged across all classes of service). Although the raw DB datasets are commonly used in academic publications (after going though some cleaning to remove frequent

flyer fares, travel by airline employees and crew, etc.), airlines generally purchase "Superset" data from the company Data Base Products (Data Base Products Inc. 2008). Superset data are a cleaned version of the DB data that are cross-validated against other data-sources to provide a more accurate estimate of market sizes. See the websites of the Bureau of Transportation Statistics (2008) or Data Base Products (2008) for additional information.

The U.S. is the only country that requires airlines to collect a 10 percent sample of used tickets. Thus, although ticketing information about domestic U.S. markets is publicly available, the same is not true for other markets. Two other sources of ticketing information include the Airlines Reporting Corporation (ARC) and the Billing Settlement Plan (BSP), the latter of which is affiliated with the International Air Transport Association (IATA). ARC is the ticketing clearinghouse for many airlines in the U.S. and essentially keeps track of purchases, refunds, and exchanges for participating airlines and travel agencies. Similarly, BSP is the primary ticketing clearinghouse for airlines and travel agencies outside the U.S.

Given an understanding of the major components of network-planning models and the OAG schedule, itinerary, and ticketing data sources that are required to support the development of these models, the next sections provide a detailed description of QSI, an alternative to logit-based market share models.

QSI Models

Market share models are used to estimate the probability a traveler selects a specific itinerary connecting an airport pair. Itineraries are the products that are ultimately purchased by passengers, and hence it is the characteristics of these itineraries that influence demand. In making their itinerary choices, travelers make tradeoffs among the characteristics that define each itinerary (e.g. departure time, equipment type(s), number of stops, route, carrier). Modeling these itinerary-level tradeoffs is essential to truly understanding air travel demand and is, therefore, one of the most important components of network-planning models.

The earliest market share models employed a demand allocation methodology referred to as QSI.[1] QSI models, developed by the U.S. government in 1957 in the era of airline regulation (Civil Aeronautics Board 1970) relate an itinerary's passenger share to its "quality" (and the quality of all other itineraries in its airport pair), where quality is defined as a function of various itinerary service attributes and their corresponding preference weights. For a given QSI model, these preference weights are obtained using statistical techniques and/or analyst intuition. Once the preference weights are obtained, the final QSI for a given itinerary is usually expressed as a linear or multiplicative function of its service

1 QSI models described in this section are based on information in the Transportation Research Board's Transportation Research E-Circular E-C040 (Transportation Research Board 2002) and on the personal experiences of Gregory Coldren and Tim Jacobs.

characteristics and preference weights. For example, suppose a given QSI model measures itinerary quality along four service characteristics (e.g. number of stops, fare, carrier, equipment type) represented by independent variables X_1, X_2, X_3, X_4 and their corresponding preference weights β_1, β_2, β_3, β_4. The QSI for itinerary *i*, QSI_i, can be expressed as

$$QSI_i = (\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4), \text{ or}$$

$$QSI_i = (\beta_1 X_1)(\beta_2 X_2)(\beta_3 X_3)(\beta_4 X_4).$$

Other functional forms for the calculation of QSI's are also possible. For itinerary *i*, its passenger share is then determined by:

$$S_i = \frac{QSI_i}{\sum_{j \in J} QSI_j}$$

where:

S_i	is the passenger share assigned to itinerary *i*,
QSI_i	is the quality of service index for itinerary *i*,
$\sum_{j \in J} QSI_j$	is the summation over all itineraries in the airport pair.

Theoretically, QSI models are problematic for two reasons. First, a distinguishing characteristic of these models is that their preference weights (or sometimes subsets of these weights) are usually obtained independently from the other preference weights in the model. Thus, QSI models do not capture interactions existing among itinerary service characteristics (e.g. elapsed itinerary trip time and equipment, elapsed itinerary trip time and number of stops). Second, QSI models are not able to measure the underlying competitive dynamic that may exist among air travel itineraries. This second inadequacy in QSI models can be seen by examining the cross-elasticity equation for the change in the passenger share of itinerary *j* due to changes in the QSI of itinerary *i*:

$$\eta^{S_j}_{QSI_i} = \frac{\partial S_j}{\partial QSI_i} \frac{QSI_i}{S_j} = -S_i QSI_i$$

The expression on the right side of the equation is not a function of *j*. That is, changing the QSI (quality) of itinerary *i* will affect the passenger share of all other itineraries in its airport pair in the same proportion. This is not realistic since, for example, if a given itinerary (linking a given airport pair) that departs in the morning improves in quality, it is likely to attract more passengers away from the other morning itineraries than the afternoon or evening itineraries.

Thus, to summarize, because QSI models have a limited ability to capture the interactions between itinerary service characteristics or the underlying competitive

dynamic among itineraries, other methodologies, such as those based on discrete choice models, have emerged in the industry.

One of the first published studies modeling air-travel itinerary share choice based on a discrete framework was published in 2003 (Coldren Koppelman Kasturirangan and Mukherjee). MNL model parameters were estimated from a single month of itineraries (January 2000) and validated on monthly flight departures in 1999 in addition to selected months in 2001 and 2002. Using market sizes from the quarterly Superset data adjusted by a monthly seasonality factor, validation was undertaken at the flight segment level for the carrier's segments. That is, the total number of forecasted passengers on each segment was obtained by summing passengers on each itinerary using the flight segment. These forecasts were compared to onboard passenger count data. Errors, defined as the mean absolute percentage deviation, were averaged across segments for regional entities and compared to predictions from the original QSI model. Regional entities are defined by time zone for each pair of continental time zones in the U.S. (e.g., East-East, East-Central, East-Mountain, East-West, …, West-West) in addition to one model for the Continental U.S. to Alaska/Hawaii and one model for Alaska/Hawaii to the Continental U.S. The MNL forecasts were consistently superior to the QSI model, with the magnitude of errors reduced on the order of 10-15 percent of the QSI errors. Further, forecasts were stable across months, including months that occurred after September 11, 2001. Additional validation details are provided in (Coldren Koppelman Kasturirangan and Mukherjee 2003).

Given an overview of the different types of itinerary choice models used in practice, the next section transitions to the modeling process used to develop logit models, using the itinerary choice problem as the foundation for the example. The discussion begins with a review of formal statistical tests used to assess the significance of individual parameters and compare different model specifications.

Model Statistics

Several statistics are used in discrete choice models to help guide the selection of a preferred model. However, although the focus of this section is on describing formal statistical tests, it is important to emphasize that the modeling process is guided by a combination of analyst intuition, business requirements, and statistics. This chapter seeks to help the reader understand how these factors are combined in practical modeling applications via a detailed example of modeling airline itinerary choices.

Formal Tests Associated with Individual Parameter Estimates

Before describing statistical tests, a brief review of statistical definitions and concepts is provided. The use of hypothesis testing is motivated by the recognition that parameter estimates are obtained from a data sample, and will vary if the

estimation is repeated on a different data sample. Stated another way, the use of hypothesis testing provides the analyst with an assessment, at a particular confidence level, that the true value for the parameter lies within the specified range. Often, the analyst is interested in knowing whether the parameter estimate is equal to a specific value (such as zero), which implies that the variable associated with the parameter does not influence choice behavior, and can be removed from the model.

Confidence intervals define a range of possible values for a parameter of a model and are directly related to the *level of uncertainty*, α. For a two-sided hypothesis test, there is a confidence of $(1 - \alpha)$ that the interval contains the true value of the parameter. High levels of uncertainty correspond to values of α that approach one (or 100 percent), whereas low levels of uncertainty correspond to values of α that approach zero (or 0 percent). Conceptually, one can loosely think of a 95 percent confidence interval in the context of a model that is estimated on 100 different (and independent) random data samples; a 95 percent confidence interval represents the range of estimated parameter values observed in (approximately) 95 out of the 100 samples.

Hypothesis testing begins with a "hypothesis," which is a claim or a statement about a property of a population, such as the population mean. The *null hypothesis*, which is typically denoted by H_0, is a statement about the value of a population parameter (such as the mean) and is designed to test the strength of the evidence against what is stated in the null hypothesis. The null hypothesis is tested directly, in the sense that the analyst assumes it is true and reaches a conclusion to either "reject H_0" or "fail to reject H_0." The *test statistic* is a value that is computed from the sample data. The test statistic is used to decide whether or not the null hypothesis should be rejected. The *critical region* is the set of all values of the test statistic that lead to the decision to reject the null hypothesis. The value that separates the critical region from the region of values where the test statistic will not be rejected is referred to as the *critical value*. The *significance level* or *level of uncertainty*, which is typically denoted by α, is the probability that the value of the test statistic will fall within the critical region, thus leading to the rejection of the null hypothesis, when the null hypothesis is true. The level of confidence is directly related to a Type I error (or false positive). A Type I error occurs the null hypothesis is rejected when in fact it is true. The level of uncertainty, α, is selected to control for this type of error. In contrast, a Type II error (false negative) occurs when the null hypothesis is not rejected, when in fact the null hypothesis is false. The probability of a Type II error is denoted by a symbol other than α to emphasize that Type II errors are not directly related to the level of uncertainty selected for the test (and will vary by problem context).

The selection of an appropriate critical value is related to the level of confidence with which the analyst wants to test the hypotheses. The selection of an appropriate significance level is somewhat arbitrary; however, in practice, it is common to use a 10 percent confidence interval (which corresponds to a critical value of 1.645 for two-sided tests) or a 5 percent confidence interval (which

corresponds to a critical value of 1.960 in two-sided tests). The relationships among the critical region, significance level, and critical values are shown in Figure 7.2.

In discrete choice modeling applications, the t-statistic is used to test a null hypothesis related to a single parameter estimate. The most common null hypothesis is that the estimate associated with the k^{th} parameter, β_k, is equal to zero:

$$H_0 : \beta_k = 0.$$

The decision rule used to evaluate the null hypothesis uses a critical value obtained from the asymptotic t distribution:

$$\text{Reject } H_0 \text{ if } \frac{\beta_k}{S_k} > \text{critical value from } t \text{ distribution}$$

where S_k is the standard error associated with the k^{th} parameter. The null hypothesis is rejected when the absolute value of the t-statistic is large. In practical modeling terms, rejection of the null hypothesis implies that the parameter estimate is different than zero, which means the variable corresponding to the parameter estimate influences choice behavior and should be retained in the

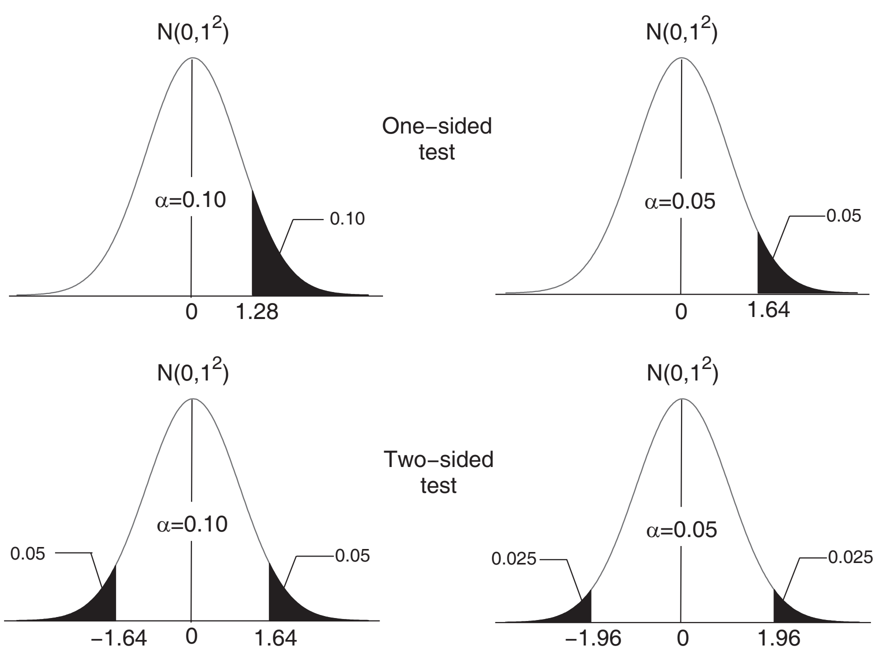


Figure 7.2 Interpretation of critical regions for a standard normal distribution

model. Failure to reject the null hypothesis occurs when the absolute values of the t-statistics are small. In practical modeling terms, the failure to reject the null hypothesis implies that the parameter estimate is close to zero, has little impact on choice behavior, and is a candidate for exclusion from the model. However, as emphasized earlier, it is important to recognize that a low t-statistic does not automatically imply exclusion of the variable from the model. Often, variables with low t-statistics are retained in the model to help support the evaluation of different policies (such as the impact of code-share agreements on market share). In addition, care should be used when excluding variables with low t-statistics early on in the modeling process, as these variables can become significant when additional variables are included in subsequent model specifications. Similarly, a large t-statistic does not automatically imply inclusion of the variable in the model (as would be the case when the sign of a parameter estimate associated with cost is positive instead of negative). These are some examples of how the modeling process is guided by statistics, analyst intuition, and business requirements.

The t-statistic is used in nested logit models to test the null hypothesis that the logsum estimate associated with the m^{th} nest, μ_m, is equal to one. Conceptually, a value close to one implies that the nesting structure is not needed, *i.e.* that the independence of irrelevant alternatives (IIA) property holds among alternatives in nest m. Formally, the null hypothesis is:

$$H_0 : \mu_m = 1$$

and the decision rule used to evaluate the null hypothesis is given as:

$$\text{Reject } H_0 \text{ if } \frac{\mu_m - 1}{S_m} > \text{critical value from } t \text{ distribution}$$

Many software packages automatically report t-statistics computed against zero, so the analyst should use caution when using t-statistics associated with logsum coefficients and ensure they are reported against one.

Formal Tests Used to Impose Linear Relationships Between Parameters

In discrete choice modeling, it is often convenient to examine whether two parameters are statistically similar to each other. For example, in itinerary choice models, the analyst may want to determine whether individuals place similar values on “small propeller aircraft” and “large propeller aircraft.” The null hypothesis is that the estimate associated with the small propeller aircraft, β_k, is equal to the parameter associated with the large propeller aircraft β_l:

$$H_0 : \beta_k = \beta_l$$

and the decision rule used to evaluate the null hypothesis is given as:

$$\text{Reject } H_0 \text{ if } \frac{\beta_k - \beta_l}{\sqrt{S_k^2 + S_l^2 - 2S_{kl}}} > \text{critical value from } t \text{ distribution}$$

where S_{kl} is the covariance associated with the estimates for the k^{th} and l^{th} parameters, and other variables are as defined earlier. Using the propeller aircraft example, rejection of the null hypothesis implies that individuals value small propeller aircraft and large propeller aircraft distinctly when making itinerary choices, and thus both variables should be retained in the model. In contrast, failing to reject the null hypothesis implies that the parameter estimates for β_k and β_l are similar, and can be combined into a single "propeller aircraft" category. Likelihood ratio tests, which are based on overall measures of model fit, can also be used to test for the appropriateness of constraining two or more parameters to be equal to each other.

Measures of Model Fit

In regression models, R^2 and adjusted R^2 measures provide information about the goodness of fit of a model. In discrete choice models, rho-squares and adjusted rho-squares play an analogous role. Conceptually, rho-squares, ρ^2, measure how much the inclusion of variables in a model improves the log likelihood function relative to a reference model. Two common reference models include an "equally likely model" and a "market shares model." In an equally likely model, each alternative in the choice set is assumed to have an equal probability of being chosen. Thus, if individual n has three alternatives in the choice set, $P_{ni} = 0.33 \ \forall \ i \in C_n$, whereas if individual q has four alternatives in the choice set, $P_{qi} = 0.25 \ \forall \ i \in C_q$. As shown in Figure 7.3, the rho-square at zero measures the improvement in log likelihoods between the estimated model, $LL(\boldsymbol{\beta})$, and the reference model, which in this case is the equally likely model, $LL(0)$. The improvement is expressed relative to the total amount of improvement that is theoretically attainable, which is the difference between the log likelihood of a perfect model $LL(*)$ and the reference model, $LL(0)$. Using the fact that the log likelihood of the perfect model is zero, ρ_0^2 is expressed as:

$$\rho_0^2 = \frac{LL(\boldsymbol{\beta}) - LL(0)}{LL(*) - LL(0)} = 1 - \frac{LL(\boldsymbol{\beta})}{LL(0)}$$

By definition, rho-squares are an index that range from zero to one. Values closer to one provide an indication that the model fits the data better.

There are several subtle, yet important points to note in Figure 7.3. First, all log likelihood values are negative. Thus, when comparing two models estimated on the same dataset the model with the "larger" (or less negative) log likelihood value fits the data better. Second, the ordering of log likelihood values shown in Figure 7.3 will hold for all models, i.e., $LL(0)$ for an equally likely model will always be less than or equal to $LL(c)$ for a constants-only model. Similarly, $LL(c)$ will always be less than or equal to $LL(\boldsymbol{\beta})$, a model that include constants and additional variables. Finally, the log likelihood of the perfect model will always be zero.

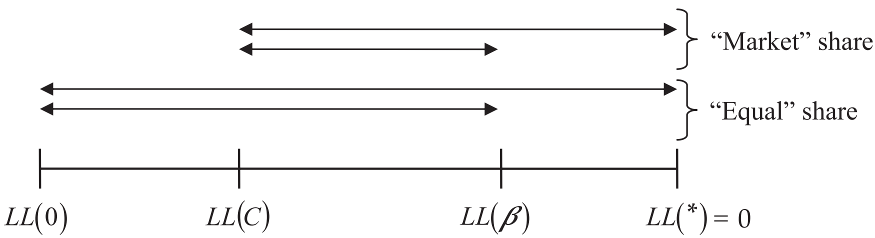


Figure 7.3 Derivation of rho-square at zero and rho-square at constants

It is appropriate to measure the goodness of fit of a model with respect to an equally likely reference model when alternative-specific constants are not included in the model. As discussed in Chapter 2, this typically occurs in situations that involve very large choice sets, such as in urban destination choice models. However, when alternative-specific constants are included in the model, it is more appropriate to measure the goodness of fit of a model with respect to a "market share" reference model, which is a model that includes a full set of identified alternative-specific constants. Conceptually, instead of assuming each alternative has an equal probability of being selected, the constants only model assumes each alternative has a probability of being selected that corresponds to the sampling shares. Thus, by using the market share model as a reference model, the improvement in log likelihood value due to including constants is excluded, and the focus shifts to measuring the improvement in model fit due to including other (and behaviorally more relevant) variables in the model. The derivation of rho-square at constants, ρ_c^2, is identical to that for ρ_0^2, except the log likelihood of the constants-only model, $LL(c)$, is used as the reference. Formally:

$$\rho_c^2 = \frac{LL(\beta) - LL(C)}{LL(*) - LL(C)} = 1 - \frac{LL(\beta)}{LL(C)}$$

One of the problems with the rho-squared measures discussed above is that they always improve when more variables are included in the model; that is, there is no penalty associated with including variables that are statistically insignificant. Adjusted rho-squares encourage parsimonious specifications by trading off the improvement in the log likelihood function against the inclusion of additional

variables. It should be noted that different formulas exist for adjusted rho-square measures. Those from Koppelman and Bhat (2006) are provided below, as they are more conservative than those reported in Ben-Akiva and Lerman (1985). The adjusted rho-squared for the zero model, $\bar{\rho}_0^2$, is given by:

$$\bar{\rho}_0^2 = \frac{LL(\boldsymbol{\beta}) - K - LL(0)}{LL(*) - LL(0)} = 1 - \frac{LL(\boldsymbol{\beta}) - K}{LL(0)}$$

where K is the number of parameters used in the model. Similarly, the adjusted rho-squared for the constants model, $\bar{\rho}_C^2$, is given by:

$$\bar{\rho}_C^2 = \frac{LL(\boldsymbol{\beta}) - K - \{LL(C) - K_{MS}\}}{LL(*) - \{LL(C) - K_{MS}\}} = 1 - \frac{LL(\boldsymbol{\beta}) - K}{LL(C) - K_{MS}}$$

where K_{MS} is the number of parameters used in the constants only model.

A second problem with rho-square measures is that they are a descriptive, and subjective, measure. Rho-squares are sensitive to the frequency of chosen alternatives in the samples. Thus, two models may be behaviorally similar, but one model may have "low" rho-squares, whereas the second may have "large" rho-squares simply due to the underlying choice frequencies. An example of this phenomenon is seen in Table 3.7, in which two no show models were estimated, one assuming the frequency of chosen alternatives reflected population rates whereas the second assumed the frequency of chosen alternatives in the sample were approximately equal. Note that ρ_0^2 drops from 0.786 to 0.129 for the dataset in which the chosen alternatives are selected in approximately equal proportions. Further, although the ρ_0^2 is much lower, the *t*-statistics associated with the parameter estimates are significant at the 0.05 level, due to the use of a more efficient estimator. This is one example of why it is difficult to use rho-square measures when evaluating the quality of a model or when comparing different model specifications. Most important, these difficulties provide a strong motivation for using the log likelihood statistics to compare different model specifications.

Tests Used to Compare Models

As discussed earlier, the *t*-statistic is used to test null hypotheses related to the value of a single parameter estimate. Likelihood ratio tests are used to compare two models. The likelihood ratio test is used when one model can be written as a restricted version of a different model. Here, "restricted" means that some parameters are set to zero and/or that one or more parameters are set equal to each other. Non-nested hypothesis tests are used when one model cannot be written as a restricted version of a second model. For example, this occurs when one model includes cost and the second model includes cost/income. Examples of how

to apply these tests are provided throughout the modeling discussion later in the chapter.

When using the likelihood ratio test, the null hypothesis is:

$$H_0 : \text{Model1 (restricted)} = \text{Model2 (unrestricted)}$$

and the decision rule used to evaluate the null hypothesis is given as:

$$\text{Reject } H_0 \text{ if } -2\left[LL_R - LL_U\right] > \text{critical value from } \chi^2_{NR,\alpha} \text{ distribution}$$

where:

LL_R is the log likelihood of the restricted model,
LL_U is the log likelihood of the unrestricted model,
NR is the number of restrictions,
α is the significance level.

By construction, the test statistic will always be positive, since the log likelihood associated with an unrestricted model will always be greater (or less negative) than the log likelihood associated with a restricted model. From a practical modeling perspective, rejecting the null hypothesis implies that the restrictions are not valid (and that the unrestricted model is preferred).

When a model cannot be written as a restricted version of another model, the non-nested hypothesis test proposed by Horowitz (1982) can be used. The null hypothesis associated with the non-nested hypothesis test is:

$$H_0 : \text{Model1 (highest } \bar{\rho}_2) = \text{Model2 (lowest } \bar{\rho}_2)$$

The decision rule, expressed in terms of the significance of the test, is:

$$\text{Reject } H_0 \text{ if } \Phi\left[-\left(-2\left(\bar{\rho}_H^2 - \bar{\rho}_L^2\right) \times LL(0) + \left(K_H - K_L\right)\right)^{\frac{1}{2}}\right] < \alpha$$

where:

$\bar{\rho}_H^2$ is the larger adjusted rho-square value,
$\bar{\rho}_L^2$ is the smaller adjusted rho-square value,
K_H is the number of parameters in the model with the larger adjusted rho-square,
K_L is the number of parameters in the model with the smaller adjusted rho-square,
Φ is the standard normal cumulative distribution function,
$LL(0)$ is the log likelihood value associated with the equally likely model.

From a practical modeling perspective, rejecting the null hypothesis implies that the two models are different (and that the model with the larger adjusted rho-square value is preferred). Failing to reject the null hypothesis implies the two models are similar.

Market Segmentation Tests

As part of the modeling process, it is typical to consider whether distinct groups of individuals exhibit different choice preferences. For example, in revenue management applications, leisure passengers are considered to be more price-sensitive, whereas business passengers are considered to be more time-sensitive. In itinerary choice applications, individuals' time of day preferences may be a function of whether they are departing or returning home. Time of day preferences may also vary as a function of the market and/or day of week, e.g. those traveling from the east coast to the west coast of the U.S. on Monday morning may have a different time of day preference than those traveling from the west coast to the east coast on Friday afternoon. Pending the availability of a sufficient sample size, estimating a model specification on different data segments (such as all EW inbound, EW outbound, WE inbound and WE outbound markets) allows the analyst to examine whether the parameter estimates are statistically different from each other (thereby reflecting different preferences across the data segments).

Assuming the same model specification is applied to each data segment, the null hypothesis is:

$$\mathrm{H}_0 : \boldsymbol{\beta}_{\text{segment 1}} = \boldsymbol{\beta}_{\text{segment 2}} = \ldots = \boldsymbol{\beta}_{\text{segment S}}$$

and the decision rule used to evaluate the null hypothesis is given as:

$$\text{Reject } H_0 \text{ if } -2\left[LL_R - \sum_{s=1}^{S} LL_s\right] > \text{critical value from } \chi^2_{NR,\alpha} \text{ distribution}$$

where:

LL_R is the log likelihood of the restricted (or pooled) model that contains all data,

LL_S is the log likelihood associated with the s^{th} data segment,

NR is the number of restrictions,

α is the significance level.

The number of restrictions in the model is defined by the following relationship:

$$NR = \sum_{s=1}^{S} K_s - K \tag{7.1}$$

where:
K_s is the number of parameter estimates in data segment s,
K is the number of parameter estimates in the pooled model.

In the case where the same specification is estimated on each data segment, the number of restrictions reduces to the following:

$$NR = K \times (S - 1) \tag{7.2}$$

In practical situations, it may be possible that some of the variables cannot be estimated within a segment in which case the less restrictive formula (Equation 7.1) applies.

Modeling Process

Data Description

Given an understanding of formal statistical tests used to assess the importance of variables and compare different model specifications, this section focuses on how to apply the formal and informal tests during the modeling process. Airline passengers' choice of itineraries is initially represented using MNL models. The analysis is based on a subset of the data used in the Coldren and Koppelman work (Coldren and Koppelman 2005a, 2005b; Coldren, Koppelman, Kasturirangan and Mukherjee 2003; Koppelman, Coldren and Parker 2008). Specifically, a single month of flight departures (January 2000) representing all airport pairs defined for two regional entities in the U.S. are represented. Regional entities are defined by time zone. In this analysis, the "East-West" regional entity contains airport pairs departing from the Eastern Time Zone and arriving in the Pacific Time Zone whereas the "West-East" regional entity contains airport pairs departing from the Pacific Time Zone and arriving on the Eastern Time Zone.

The data used for the analysis is from three primary sources. CRS (or MIDT) booking data contain information on booked itineraries across multiple carriers. As stated in Coldren, Koppelman, Kasturirangan, and Mukherjee (2003) "CRS data are commercially available and compiled from several computer reservation systems including Apollo, Sabre, Galileo, and WorldSpan as well as Internet travel sites such as Orbitz, Travelocity, Expedia, and Priceline. The CRS data are believed to include 90 percent of all bookings during the study period. However, increasing use of direct carrier and other Internet booking systems has reduced the proportion of bookings reported by this source, a problem that will have to be addressed in the foreseeable future." In addition to providing information on the itinerary origin and destination and the number of individuals traveling together on the same booking record, CRS data provide detailed information for each flight leg in the itinerary. For each leg, CRS data contain its origin and destination, flight

number, departure and arrival dates, departure and arrival times, and marketing and operating carrier(s). By definition, a marketing carrier is the airline who sells the ticket, whereas the operating carrier is the airline who physically operates the flight. For example, a code-share flight between Delta and Continental could be sold either under a Delta flight number or a Continental flight number. However, only one plane is flown by either Delta or Continental.

The other two data sources used in the analysis are from the Official Airline Guide (OAG) and Superset (OAG Worldwide Limited 2008; Data Base Products Inc. 2008) OAG contains leg-based information on the origin, destination, flight number, departure and arrival times, days of operation, leg mileage, flight time, operating airline, and code-share airline (if a code-share leg). Superset data, described in detail earlier in this chapter in the overview of major components of network-planning models, provide information on quarterly airport-pair average fares averaged across all classes of service and times of day for each airline serving the airport-pair.

Table 7.1 provides definitions for variables explored during the modeling process. Several of these variables merit further discussion. With respect to level of service, two formulations will be explored in the modeling process. The first formulation represents level of service simply as the (average) value passengers associated with a non-stop, direct, single-connect, or double-connect itinerary. The second formulation represents level of service with *respect to the best level of service* available in the airport pair and reflects the analyst's intuition that an itinerary with a double-connection is much more onerous to passengers when the best level of service in the market is a non-stop than when the best level of service in the market is a single-connection.

Two formulations to represent passengers' preferences for time of day are also explored in the modeling process. In the first formulation, preferences for departure times are represented via the inclusion of time of day dummy variables for each hour of the day. In the second formulation, the dummy variables are replaced by six sine and cosine functions, which create a continuous distribution representing time of day preferences. Finally, it is important to note that in the major carrier's MNL itinerary share model, preferences for departure times are represented via the inclusion of time of day dummy variables for each hour of the day. In practice, there are other methods based on schedule delay formulations that are currently in use. Unfortunately, the terminology that has been used to describe the schedule delay functions is often referred to as a "nested logit model" within the airline community, which is incorrect. To clarify, a schedule delay function captures the difference between an individual's expressed departure time preference and the actual departure time of a flight, whereas a "nested logit model" refers to the NL probability expression derived in Chapter 3.

Another common industry practice reflected in itinerary share models is to include carrier presence variables. Numerous studies have found that increased carrier presence in a market leads to increased market share for that carrier (Algers and Beser 2001; Nako 1992; Proussaloglou and Koppelman 1999; Suzuki

Table 7.1 Variable definitions

Variable	Description
Fare ratio	Carrier average fare divided by the industry average fare for the airport-pair multiplied by 100.
Carrier	Dummy variable representing major US domestic carriers. “All other” (non-major) carriers are combined together in a single category.
Level of service	Dummy variable representing the level of service of the itinerary (nonstop, direct, single-connect, double-connect). Level of service is measured in some models with respect to the best level of service available in the airport-pair.
Time of day—discrete	Dummy variable for each hour of the day (based on the local departure time of the first leg of the itinerary).
Time of day—continuous	Three sine and three cosine waves are used to represent itinerary departure time. For example, sin 2PI = sin {(2PI*departure time)/1440} where departure time is expressed as minutes past midnight. Frequencies are for 2PI, 4PI, and 6PI.
Point of sale weighted presence	Point of sale weighted presence of carrier at the origin and destination airports. Presence is measured as the percentage of operating departures out of an airport, including connection carriers. Point of sale weighted presence is an integer between 0 and 100 and is used in models that predict itinerary choice for all departing and returning passengers.
Origin presence	Presence at the airport at the origin of an itinerary that is an integer between 0 and 100. Used when modeling itinerary choice of outbound/departing passengers.
Destination presence	Presence at the airport at the destination of an itinerary that is an integer between 0 and 100. Used when modeling itinerary choice of inbound/returning passengers.
Code share	Dummy variable indicating whether any leg of the itinerary was booked as a code share. Code share is represented in some models as a function of airline presence, i.e., a “small code share” reflects an itinerary that operates in a market where the airline has a small operating presence, (specifically a presence score of 0-4) while a “large code share” represents a market with a presence score of 5 or higher.
Propeller aircraft	Dummy variable indicating whether the smallest aircraft on any part of the itinerary is a propeller aircraft. In some models, this is further broken down into “small prop” and “large prop”.
Regional jet	Dummy variable indicating whether the smallest aircraft on any part of the itinerary is a regional jet aircraft. In some models, this is further broken down into “small RJ” and “large RJ”.
Commuter	Dummy variable indicating whether the smallest aircraft on any part of the itinerary is a propeller or a regional jet aircraft.
Narrow-body	Dummy variable indicating whether the smallest aircraft on any part of the itinerary is a narrow-body aircraft.
Wide-body	Dummy variable indicating whether the smallest aircraft on any part of the itinerary is a wide-body aircraft.

Tyworth and Novack 2001). In this modeling process, a "point of sale weighted airport presence" variable is used to represent carrier presence at both the origin and destination. Similarly, an origin (destination) presence variable represents carrier presence at the origin (destination) airport. By definition, a simple round trip ticket contains two itineraries: a departing itinerary, which represents the outbound portion of a trip and a returning itinerary, which represents the inbound portion of a trip. When separate models are estimated for outbound and inbound passengers, market presence at the individuals' home locations can be modeled. This is done by using the origin presence of an itinerary for departing passengers and the destination presence of an itinerary for returning passengers.

As a final note, it is often desirable from a business perspective to be able to differentiate impacts associated with adding a code-share flight in an airport pair that has a strong operating presence by the marketing carrier versus in an airport pair in which the marketing carrier has a weak operating presence. That is, one expects the effect of a code-share to be larger in markets in which the marketing carrier has a stronger operating presence. For example, assume that United operates a flight between Chicago O'Hare (ORD) and Paris Charles de Gaulle (CDG) international airports, and that United is debating whether to pursue a code-share agreement with British Airways or Air France (i.e., which airline to select as the marketing carrier). In this example, Air France should be selected as the code-share partner, since Air France has a higher operating presence in the ORD-CDG market. That is, potential customers are more likely to recognize the Air France brand than the British Airways brand, resulting in Air France being a better marketing (or code-share) partner.

Descriptive Statistics

Before launching into model estimation, it is very helpful (and highly recommended!) that the analyst become familiar with the data. Descriptive statistics can help detect subtle errors that may have occurred when creating the estimation dataset. They are also useful in diagnosing estimation problems (such as lack of convergence, lack of *t*-statistics, etc.) These types of estimation problems can occur when the sample size associated with a variable included in the model specification is low. One of the most common errors the authors have seen students make when using real-world datasets relates to misinterpretation and/or failure to understand how missing values are coded. That is, it is important to recognize that in some datasets, a value of zero physically means "zero" whereas in other datasets, missing values can be coded as zero or a number typically outside the reasonable range associated with a variable, e.g., if an individual's age can take on values from 0 to 99, a missing value could be represented as -1 or 999. Through examining descriptive statistics (such as the mean, minimum, and maximum values associated with a variable, along with other measures of location and dispersion), these and other coding problems can often be detected.

Selective descriptive statistics are described for the EW dataset (similar results apply to the WE dataset). The EW dataset contains 12,681 choice sets (each representing a unique airport pair day of week). The mean number of itineraries (or alternatives) available in a choice set is 78.8, with a standard deviation of 56.6, indicating that there is a large variation that the number of available itineraries can have across choice sets. The number of available itineraries ranges from a minimum of one to a maximum of 313.

The distribution of available itineraries in the EW market is shown in Table 7.2. Although there are more than 3,000 weekly non-stop flights, non-stops represent only 0.55 percent of all available itineraries; there are many more single-connections (42 percent) and double-connections (57 percent) created by the airline's itinerary generation algorithm. Further, although non-stops and directs combined represent 1 percent of all available itineraries, they carry 7.2 percent of all booked passengers. Most passengers (89.9 percent) book single-connections in the EW market, and very few (2.9 percent) book double-connections.

One question that naturally arises from the above discussion is whether 89.9 percent of passengers are booking single-connections because they prefer them or because they are the best option available (e.g., there is no non-stop or direct service in the airport pair). Table 7.3 shows the distributions of available itineraries and booked passengers with respect to the best level of service in

Table 7.2 Descriptive statistics for level of service in EW markets (all passengers)

	# itineraries available	% itineraries	# booked passengers	% booked passengers
Non-stop	3,046	0.55%	6,005	4.4%
Direct	2,824	0.51%	3,826	2.8%
Single-Connection	233,584	41.98%	124,146	89.9%
Double-Connection	317,000	56.97%	4,056	2.9%
TOTAL	556,454		138,033	

the market. Thus, although 89.9 percent of passengers book single-connections, approximately half of these occur in markets in which the best level of service available to the passenger is a single-connection. In addition, 20 percent of all bookings occur on single-connecting itineraries when the best level of service is a direct itinerary. A direct itinerary is similar to a single-connection in that it involves a stop. However, distinct from a single-connection itinerary, the flight number associated with both legs of the direct itinerary are the same and (usually) passengers do not change equipment at the stop over location. Only 22 percent of all bookings occur on single-connection itineraries in markets where the best

level of service is a non-stop market. Table 7.3 also reveals that few passengers choose double-connections in markets in which the best level of service is a non-stop or direct. Based on the analysis of descriptive statistics for level of service, the analyst may decide to estimate two different models. The first includes the average preference associated with a non-stop, direct, single-connection, and double-connection whereas the second captures interactions in these preferences with respect to the best level of service in the market. This is one example of how the use of descriptive statistics can help guide the analyst in deciding which variables to include in a model and how descriptive statistics can be used in the "pre-modeling" stage.

Interpretation of Dependent Variable

There is one characteristic of the airline itinerary data that needs to be discussed further before illustrating the process of specifying and refining a MNL model.

Table 7.3 Descriptive statistics for level of service with respect to best level of service in EW markets (all passengers)

	# itineraries available	% itineraries	# booked passengers	% booked passengers
NS in NS	3,046	0.55%	6,005	4.35%
Direct in NS	1,062	0.19%	1,438	1.04%
SC in NS	67,315	12.10%	29,791	21.58%
DC in NS	35,761	6.43%	78	0.06%
Direct in Direct	1,762	0.32%	2,388	1.73%
SC in Direct	41,428	7.44%	27,622	20.01%
DC in Direct	38,814	6.98%	228	0.17%
SC in SC	124,841	22.44%	66,733	48.35%
DC in SC	217,483	39.08%	1,953	1.41%
DC in DC	24,942	4.48%	1,797	1.30%
TOTAL	556,454		138,033	

Key: NS = nonstop; SC = single connection; DC = double connection

The EW dataset contains 12,681 observations. An observation is defined as a unique airport pair day of week (e.g., all itineraries between Boston-San Francisco on a representative Monday in January, 2000). The dependent variable represents the number of passengers that choose each itinerary. Thus, within an observation, the total number of "observed choices" or passengers can be greater than one. Indeed, as shown in Table 7.3, the number of booked passengers in the EW dataset is 138,033, representing an average of 10.9 passengers per observation.

Conceptually, the itinerary choice data do not represent true "disaggregate" passenger choices in the sense that the choice scenario is not customized to each passenger (i.e., average fares are used and it is assumed itineraries are always available). The itinerary choice data represent "aggregate" passenger choices in the sense that we know for each airport pair day of week, the total number of passengers choosing each itinerary. Stated another way, the decision unit of analysis is "airport pair day of week." From a statistical perspective, if the number of passengers is used as the dependent variable, the significance of *t*-statistics will be inflated, implying variables are more significant than they "really" are. To correct for this bias, a weighted dependent variable can be used. The weight is selected so that the sum of the dependent variables over all observations is equal to the total number of observations. For example, given the WE dataset with EW dataset contain 12,681 unique observations (or airport day of week choice sets) and 138,033 bookings, a weight of 12,681/138,033 = 0.0919 would be used. Weighting also has the advantage of decreasing the time it takes to estimate a model. Models reported in this section were estimated using Gauss (Aptech Systems Inc. 2008). A comparison of running times for Model 4 was 2.55 minutes when using the weighted dependent variable and 10.23 minutes for using the unweighted dependent variable. Absolute running times are not important (as code has not been optimized for speed). What is important to note is the difference in running time between the weighted and unweighted dependent variables. In this case, using the weighted dependent variable decreases estimation time by a factor of four.

Depending on the software used to estimate logit models, re-scaling the independent variables may also help decrease estimation time, particularly when the parameter estimates differ by several orders of magnitude. That is, at a fundamental level, solving for the parameters of a MNL (NL) model is a linear (non-linear) program. As such, many of the principles or "tricks" that are used in the linear and non-linear optimization should apply to the solution of parameters from a discrete choice model. However, among the discrete choice modeling community, little attention has been placed on developing more efficient algorithms and data storage schemes for these applications. Due to the substantially larger datasets encountered in air travel applications (versus the more traditional transportation, marketing, and economics applications), it would not be surprising if the needs of the airline community spurred new methodological developments in this area.

Base MNL Models

Different approaches can be used to select a preferred model specification. One approach is to start with a simple model that includes the variables the analyst believes are "most important" to the choice process. For example, in itinerary choice applications, it is common to include four key variables in a model: fare, level of service, departure time, and carrier. Several models can be estimated to explore how the inclusion of these key variables influences model fit. That is, at this stage, the analyst compares specifications that use linear versus non-linear representations, discrete versus continuous representations, different groupings for categorical variables, etc. For example, a model that includes log of fare may fit the data better than a model that includes fare. Even if the latter specification fits the data better, the analyst may decide to use the log of fare because it has a stronger behavioral foundation, i.e., the use of log(fare) captures the analyst's belief that a \$50 increase in a \$100 fare will have a larger impact on passenger choice than a \$50 increase in a \$1,000 fare. This is one example of how the selection of a preferred model specification is guided by both statistics and behavioral theories. After initial model specifications with key variables has been included, additional variables (thought to be less important in influencing choice and/or that have small sample sizes) can be incorporated in more advanced specifications.

The modeling process described above is an "incremental" approach in the sense that the analyst begins with a simple model specification and incrementally adds variables to obtain a more complex specification. This approach is recommended for novice modelers and those new to discrete choice modeling. This is because by comparing different model specifications, the analyst can more easily detect the presence of multi-collinearity and more easily isolate underlying causes of estimation problems (such as failure to converge or lack of *t*-statistics due to a miscoded variable, unstable *t*-statistics associated with a variable with a low sample size, etc.). An alternative approach is to start with a complex model specification and delete variables that are not significant. The first approach is demonstrated in this section.

Table 7.4 shows the results of four MNL model specifications that include variables for carrier, fare, level of service, and time of day (the key variables the analyst believes have the strongest impact on itinerary choice). Nine carriers are represented in the data: Air Canada, American, America West, Continental, Delta, Northwest, United, US Airways, and "all others." The "all other" category contains airlines that (each) have less than a 5 percent share across the EW markets. The coefficients associated with these airlines are not shown in any of the model specifications for confidentiality reasons. That is, carrier constants are suppressed because they are a reflection of the strength of a carrier's brand (and indirectly capture the strength of frequent flyer programs, advertising, etc.).

The second variable is fare ratio, which is derived from the Superset data. Specifically, Superset data contain information on the average fare sold by each carrier in an airport pair. This fare is very "aggregate" or "high-level" in the

Table 7.4 Base model specifications for EW outbound models

	MNL 1: Base Model	MNL 2: LOS	MNL 3: Time of Day	MNL 4: LOS and TOD
Carrier Attributes				
Fare ratio	-0.0006 (0.7)	-0.0006 (0.7)	-0.0005 (0.1)	-0.0005 (0.6)
Carrier constants (proprietary)	--	--	--	--
Level of Service				
Non-stop (reference)	0		0	
Direct	-2.75 (44.2)		-2.73 (44.3)	
Single-connect	-4.43 (144)		-4.43 (146)	
Double-connect	-9.44 (18.6)		-9.43 (18.6)	
Level of Service w.r.t. Best Level of Service				
Non-stop in Non-stop (ref.)		0		0
Direct in Non-stop		-2.72 (33)		-2.70 (33)
Single-Connect in Non-stop		-4.44 (135)		-4.43 (136)
Double-Connect in Non-stop		-10.36 (4.5)		-10.35 (4.5)
Direct in Direct (ref.)		0		0
Single-Connect in Direct		-1.66 (18)		-1.66 (18)
Double-Connect in Direct		-7.17 (4.2)		-7.18 (4.2)
Single-Connect in Single-Connect (ref.)		0		0
Double-Connect in Single-Connect		-4.82 (8.8)		-4.82 (8.8)
Categorical Time of Day Formulation				
5–6 A.M.	-0.231 (0.8)	-0.231 (0.8)		
6–7 A.M. (ref.)	0	0		
7–8 A.M.	0.212 (4.4)	0.213 (4.5)		
8–9 A.M.	0.228 (4.7)	0.229 (4.7)		
9–10 A.M.	0.285 (6.1)	0.286 (6.1)		

Table 7.4 *Concluded*

	MNL 1: Base Model	MNL 2: LOS	MNL 3: Time of Day	MNL 4: LOS and TOD
10–11 A.M.	0.271 (4.4)	0.271 (4.4)		
11–12 noon	0.019 (0.3)	0.018 (0.3)		
12–1 P.M.	-0.036 (0.7)	-0.036 (0.7)		
1–2 P.M.	-0.187 (2.4)	-0.187 (2.4)		
2–3 P.M.	-0.298 (4.4)	-0.298 (4.4)		
3–4 P.M.	-0.251 (4.3)	-0.251 (4.3)		
4–5 P.M.	-0.345 (5.3)	-0.344 (5.3)		
5–6 P.M.	-0.362 (6.8)	-0.361 (6.8)		
6–7 P.M.	-0.232 (4.3)	-0.233 (4.3)		
7–8 P.M.	-0.220 (4.2)	-0.220 (4.2)		
8–9 P.M.	-0.525 (10.1)	-0.525 (10.1)		
9–10 P.M.	-0.715 (10.8)	-0.715 (10.8)		
10–Midnight	-0.920 (12.0)	-0.920 (12.0)		
Continuous Time of Day Formulation				
Sin 2pi			0.058 (1.4)	0.057 (1.4)
Sin 4pi			-0.283 (6.6)	-0.284 (6.6)
Sin 6pi			-0.040 (1.5)	-0.040 (1.5)
Cos 2pi			-0.624 (11.0)	-0.625 (11.0)
Cos 4pi			-0.247 (10.7)	-0.247 (10.8)
Cos 6pi			-0.047 (2.8)	-0.047 (2.8)
Model Fit Statistics				
LL at zero	-59906.83	-59906.83	-59906.83	-59906.83
LL at convergence	-37447.54	-37444.25	-37456.22	-37452.84
Rho-square w.r.t. zero	0.3749	0.3750	0.3748	0.3748
# parameters/adj. rho-square zero	29 / 0.3744	32 / 0.3744	18 / 0.3745	21 / 0.3745

Key: LOS = level of service; TOD = time of day. See Table 7.1 for variable definitions. Carrier constants suppressed for confidentiality reasons.

sense that it represents an average over all classes of service and times of day for all itineraries departing in over a three-month time frame. Disaggregate fare information representing fares purchased on a specific itinerary were not available for the analysis. Because fares differ by length of haul and across airport pairs represented in the dataset, a "fare ratio," defined as the carrier average fare (for the quarter) divided by the industry average fare for the airport pair multiplied by 100 was used in the analysis. A fare ratio greater than one indicates that the carrier sold fares higher than market average whereas a fare ratio less than one indicates the carrier sold fares lower than market average. Intuitively, the coefficient associated with fare ratio should be negative to reflect passenger preferences for itineraries with lower fares. This is observed in all four models in Table 7.4. However, the parameter is not significant at the 0.05 level, as observed from the *t*-stats below 2.0. Due to the perceived importance of this variable in influencing choice of itinerary, the variable is retained throughout subsequent model specifications. That is, we do not want to drop a variable "too soon" in the modeling process, as its parameter estimate may become significant when additional variables are included in the model.

The third variable is level of service. Two representations are examined: the first (shown in Models 1 and 3) represents level of service using three parameters: direct, single-connection, and double-connection. Intuitively, since non-stop itineraries are defined as the reference category, the coefficients associated with the level of service variables should be negative. This is observed in both Models 1 and 3, which show a clear preference of passengers for non-stop itineraries (followed by directs, single-connects, and double-connects). In addition, all parameter estimates are significant at the 0.05 level. The second formulation (shown in Models 2 and 4) represents level of service with respect to the best level of service. Note that since only differences in utility (within a choice set) are uniquely identified, several references must be defined. That is, when the best level of service in a market (or choice set) is a non-stop, parameters for directs, single-connections, and double-connections can be estimated. Similarly, when the best level of service in a market is a direct, only two parameters can be estimated. Setting directs as the reference, parameters for single-connections and double-connections can be estimated. Similar logic applies to the fact that only one parameter (for double-connections) can be estimated for choice sets in which the best level of service in the market is a single-connection.

The results of this formulation are shown in Models 2 and 4. Because the reference is defined as the "best" level of service within each case, all level of service parameters are expected to be negative. A comparison of the relative magnitudes across the best level of service in the market shows that double-connections are much more onerous in markets in which the best level of service is a non-stop (-10.4) than in markets in which the best level of service is a direct (-7.2) or single-connection (-4.8). Similarly, single-connections are much more onerous in non-stop markets (-4.4) than in direct markets (-1.7). All level of service parameter estimates are significant at the 0.05 level.

The likelihood ratio test can be used to evaluate whether the improvement in log likelihood associated with using three additional parameters to represent level of service with respect to the best level of service in the market is statistically significant. Formally, the null hypothesis is:

$$H_0 : \beta_{\text{Single Cnx in NS}} = \beta_{\textit{Single Cnx in Dir}}$$

$$\beta_{\textit{Double Cnx in NS}} = \beta_{\textit{Double Cnx in Dir}} = \beta_{\textit{Double Cnx in Single Cnx}}$$

And the corresponding decision rule is:

Reject H_0 if $-2\left[-37447.54 - (-37444.25)\right] > \chi^2_{3,\,0.05}$

Reject H_0 if $6.58 > 7.81$

In this case, the null hypothesis cannot be rejected at the 0.05 level. (However, it can be rejected at the 0.10 level.) Note that this is *in spite of* the fact that all *t*-statistics associated with the level of service variables are significant at the 0.05 level. From a practical perspective, the results of the likelihood ratio test imply that the two formulations for level of service are equivalent, and that the simpler model specification of Model 1 should be used. However, given the stronger behavioral foundation of the second formulation combined with the fact the null hypothesis can be rejected at the 0.10 level, the formulation with respect to the best level of service in the market is retained for further model exploration.

The fourth variable is time of day. Two formulations are used to represent passengers' departure time preferences. The first formulation (shown in Models 1 and 2) uses dummy variables for each departure hour. Due to small sample sizes, flights departing from 10 PM to midnight are combined into a single category. Because no flights depart from midnight to 5 AM and few flights depart from 5 AM to 6 AM, the reference category of 6 to 7 AM is used in the analysis. The second formulation replaces the categorical time of day specification with a continuous specification that combines three sine and three cosine functions. For example, sin 2PI is represented as:

$$\text{sin 2PI} = \sin\left\{\left(2\pi \times \text{departure time}\right)/1440\right\}$$

where departure time is expressed as minutes past midnight. Frequencies of 2PI, 4PI, and 6PI are used in the continuous specification. The results of this specification are shown in Models 3 and 4. As a side note, Carrier (2008) proposed a modification to this formulation to account for cycle lengths that are shorter than 24 hours. Formally, the equation $\beta_1 \sin\,(2\pi h/1440) + \beta_2 \text{cosin}\,(2\pi h/1440) + \ldots$ is replaced with:

$$\beta_1 \sin\,\{2\pi\,(h - s)/\,d\} + \beta_2 \text{cosin}\,\{2\pi\,(h - s)/\,d\} + \ldots$$

where:

$1 - e \leq d \leq 24$ and $0 \leq s \leq e$

where *e* and 1 represent the departure times of the earliest and latest itineraries in the market, respectively, *h* represents the departure time, *s* represents the start time of the cycle (which is not uniquely identified and can be set to an arbitrary value) and *d* represents the cycle duration. The examples in this chapter use the 24-hour period, as Carrier's formulation leads to a nonlinear-in-parameters function, which he solved using a trial-and-error method. The trial-and-error method (often used by discrete choice modelers when they encounter nonlinear-in-parameters functions) essentially fixes *d* to different values and estimates the remaining parameters. The value of *d* that results in the best log likelihood value is the preferred model.

The interpretation of time of day parameter estimates from the discrete and continuous formulations is shown in Figures 7.4 and 7.5. Both formulations show passengers prefer itineraries departing early in the morning or later in the afternoon. Intuitively, this makes sense as departing passengers may want to leave early whereas returning passengers may want to leave later in the day. One of the problems with the discrete formulation is that counter-intuitive results can occur in what-if scenarios when analysts make slight changes in the timing of itineraries. For example, an itinerary whose departure time moves from 10:59 AM to 11:01 AM has a change in utility from 0.27 to 0.02 (indicating that the 11:01 departure time is "much" less preferred than the 10:59 AM departure). This problem is mitigated by the use of the continuous formulations.

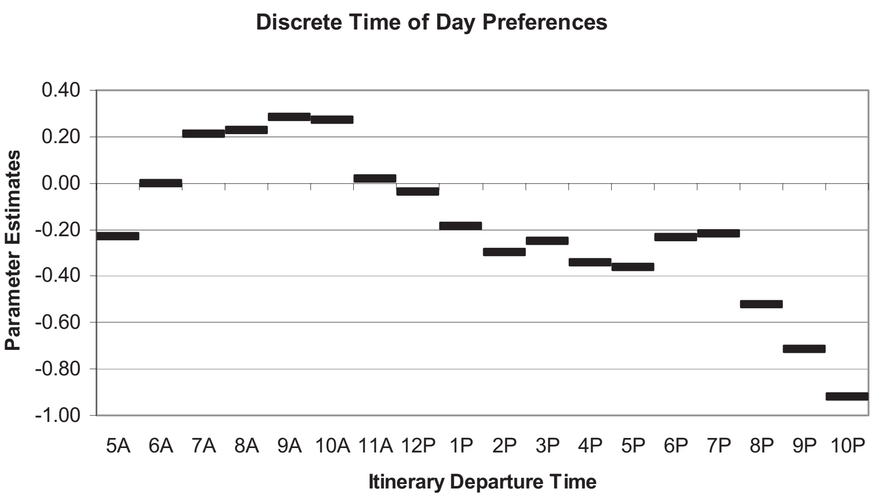


Figure 7.4 Interpretation of time of day from MNL model 2

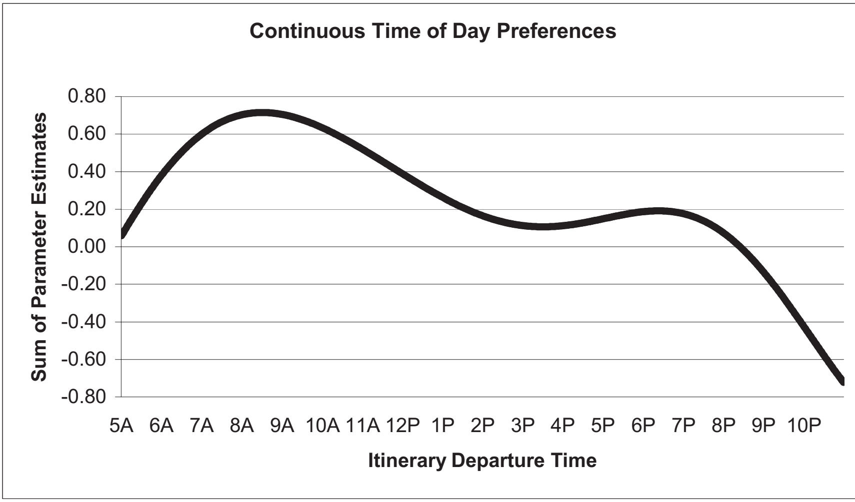


Figure 7.5 Interpretation of time of day from MNL model 4

Several of the parameter estimates for both the discrete and continuous time of day parameters are not significant at the 0.05 level. In the case of the discrete formulation, this occurs when the parameter estimates are close to the zero intercept (which implies the preference for the 11 AM–1 PM departures is similar to that of the reference category, or 6 AM–7 AM departures). In the case of the continuous time of day specification, the amplitude associated with the sin 2PI and sin 6PI frequencies are small, indicating that these frequencies do not contribute to the overall shape of the curve (and may be dropped from the specification). For now, we will retain all frequencies in the model to facilitate comparisons among models for the EW and WE markets.

The non-nested hypothesis test can be used to statistically compare the continuous and discrete time of day formulations (e.g., to compare Models 1 and 3). Formally, the null hypothesis is:

$$H_0 : \text{Model 3 (highest } \bar{\rho}_2) = \text{Model 1 (lowest } \bar{\rho}_2)$$

and the significance of the test (including the suppressed carrier constants) is given as:

$$\Phi\left[-\left(-2\left(\bar{\rho}_H^2 - \bar{\rho}_L^2\right)\times LL(0) + \left(K_H - K_L\right)\right)^{\frac{1}{2}}\right]$$

$$\Phi\left[-\left(-2\left(0.3745 - 0.3744\right)\times -59906.83 + \left(18 - 29\right)\right)^{0.5}\right]$$

$$\Phi\left(\text{-}0.99\right) = 0.081$$

From a practical perspective, the significance of the test implies that the two formulations for time of day are statistically equivalent when a significance of 0.05 is used, but that the discrete time of day formulation is preferred when a significance of 0.10 is used. The log likelihood values for these two formulations are very similar, and given the stronger behavioral foundation, in addition to the forecasting advantages, the continuous time of day formulation is retained as the preferred model specification.

Models 1 through 4 viewed together are a reflection of the "incremental" modeling approach that explores isolated and joint impacts of using different formulations for level of service and time of day. Model 1 represents the "simplest" level of service formulation in combination with dummy variables for time of day. Examining the time of day results from this first model enable the analyst to see which continuous formulations (such as the sine and cosine functions) may be appropriate alternatives. Model 2 examines only the impact of relaxing the level of service representation, whereas Model 3 examines only the impact of using the continuous time of day representation. Model 4 looks at the impact of both of the representations. Relevant model comparison tests, summarized in Table 7.5, confirm the results discussed above.

Table 7.5 Formal statistical tests comparing models 1 through 4

	Model 2	**Model 3**	**Model 4**
LRT to reject Model 1	6.6, 3, 7.8, 0.087	NA	NA
NNT to reject Model 1	NA	0.081	NA
NNT to reject Model 2	NA	NA	0.081
LRT to reject Model 3	NA	NA	6.8, 3, 7.8, 0.079

Key: LRT = Likelihood Ratio test; NNT = Non-nested hypothesis test. Information provided for LRT = Likelihood Ratio statistic, degrees of freedom, critical value, rejection significance level. Information provided for NNT= rejection significance level.

Model 4 is used as the "base" model on which additional variables are included. Models 5 and 6, shown in Table 7.6, look at the impact of carrier presence and code-shares on itinerary choice. Origin presence measures a carrier's presence at an airport. Origin presence is used for departing itineraries to reflect the carrier's presence at the passenger's home location (i.e., where the passenger is assumed to reside). Intuitively, it is expected that a large presence will result in proportionately more market share for the carrier. In the airline industry, this effect is sometimes referred to as the "halo effect." For example, assume a carrier controls 70 percent of all departures out of an airport. The "halo" effect refers to the fact that more than 70 percent of passengers departing from that airport tend to chose that carrier, due to effects of local advertising, desire to support the "hometown" airline, greater ability of passengers to concentrate frequent flyer miles on the hometown airline,

etc. Consistent with this logic, the parameter estimate associated with carrier presence is positive.

Model 5 also contains a code-share dummy variable that indicates whether any leg of the itinerary was booked as a code-share. Conceptually, a code-share itinerary is not expected to draw as many passengers as its equivalent non-code-share itinerary. That is, a code-share flight refers to a flight that is "marketed" by one airline, but operated by a different airline. In this case, the operating carrier of the first leg of the itinerary is generally responsible for check-in procedures. Thus, to avoid passenger confusion (i.e., a "ticket" that shows the marketing carrier airline with instructions to check-in with the operating carrier), travel agents may book the itinerary on the operating carrier. The parameter associated with code-share itineraries is negative and large in relative magnitude, indicating that itineraries marketed and operated by the same carrier are preferred to those marketed by one carrier and operated by a different carrier.

When evaluating which flights are good candidates for code-share agreements, it is often helpful for an airline to distinguish between code-shares offered in markets where they have a strong vs. weak presence. Model 6 incorporates this effect and shows that an airline considering a code-share flight will perform better in markets where the marketing carrier partner is stronger than markets in which the marketing carrier partner is weaker.

Likelihood ratio tests (shown at the bottom of Table 7.6) clearly reject the null hypotheses that Model 5 = Model 4 and that Model 6 = Model 5. Thus, Model 6, which includes carrier presence and code-share factors differentiated by whether the operating carrier has a large or small market presence, is used as the new base model for exploring the effects of adding equipment type.

Model 7 examines the impact of six equipment types on itinerary choices: small propeller, large propeller, small regional jet, large regional jet, narrow-body, and wide-body. Here, equipment type refers to the smallest equipment type on the itinerary. Since the largest equipment type is the reference, parameter estimates are expected to be negative to reflect passengers' preferences to fly on larger planes. The parameters in Model 7 are all negative, but the relative magnitudes are not consistent with expectations. That is, both small and large propeller flights are expected to be more onerous than small and large regional jets. Thus, although the likelihood ratio test rejects the null hypothesis that Model 6 = Model 7, suggesting that equipment type does influence itinerary choices, this is not a model that would be appropriate to use in forecasting, as it would give counter-intuitive results.

Model 8 eliminates the small and large distinctions in propeller and regional jets. The parameter estimate for propellers (-1.07) lies between those observed in Model 7 for small and large propellers (-1.10 and -0.99). Similarly, the parameter estimate for regional jets (-0.99) falls between those observed for small and large regional jets (-1.09 and -0.21), and in this case is closer to the value that has the larger *t*-statistic. This is a common pattern that is often observed when combining categories. However, the pattern will not always be observed, as parameters are simultaneously estimated. In the case where variables are highly correlated, adding

Table 7.6 Equipment and code-share refinement for EW outbound models

	MNL 5: Code Share	MNL 6: Code Share 2	MNL 7: Equip 1	MNL 8: Equip 2	MNL 9: Equip 3
Carrier Attributes					
Fare ratio	-0.0055 (5.6)	-0.0057 (5.8)	-0.0063 (6.3)	-0.0063 (6.3)	-0.0063 (6.3)
Carrier constants (proprietary)	--	--	--	--	--
Level of Service					
Non-stop in Non-stop	0	0	0	0	0
Direct in Non-stop	-2.59 (32)	-2.58 (32)	-2.57 (31)	-2.57 (31)	-2.57 (32)
Single-Connect in Non-stop	-4.17 (117)	-4.16 (117)	-4.08 (114)	-4.09 (114)	-4.09 (115)
Double-Connect in Non-stop	-9.87 (4.3)	-9.85 (4.3)	-9.52 (4.1)	-9.54 (4.1)	-9.54 (4.1)
Direct in Direct	0	0	0	0	0
Single-Connect in Direct	-1.59 (17)	-1.59 (17)	-1.51 (16)	-1.51 (16)	-1.51 (16)
Double-Connect in Direct	-6.91 (4.0)	-6.90 (4.0)	-6.59 (3.8)	-6.60 (3.8)	-6.60 (3.8)
Single-Connect in Single-Connect	0	0	0	0	0
Double-Connect in Single-Connect	-4.60 (8.4)	-4.59 (8.4)	-4.39 (6.0)	-4.40 (8.0)	-4.41 (8.0)
Time of Day					
Sin 2pi	0.059 (1.5)	0.060 (1.5)	0.044 (1.1)	0.046 (1.1)	0.046 (1.1)
Sin 4pi	-0.291 (6.7)	-0.290 (6.7)	-0.292 (6.7)	-0.291 (6.7)	-0.291 (6.7)
Sin 6pi	-0.047 (1.8)	-0.048 (1.8)	-0.059 (2.2)	-0.057 (2.2)	-0.057 (2.2)
Cos 2pi	-0.630 (11)	-0.630 (11)	-0.637 (11)	-0.633 (11)	-0.634 (11)
Cos 4pi	-0.264 (12)	-0.264 (12)	-0.249 (11)	-0.247 (11)	-0.247 (11)
Cos 6pi	-0.046 (2.7)	-0.046 (2.7)	-0.049 (2.9)	-0.047 (2.8)	-0.048 (2.8)
Aircraft Type					
Small prop			-1.10 (5.0)		
Large prop			-0.99 (3.5)		
Propeller aircraft			--	-1.07 (5.8)	
Small regional jet			-1.09 (7.0)		
Large regional jet			-0.21 (0.6)		
Regional jet			--	-0.99 (6.8)	

Table 7.6 Concluded

	MNL 5: Code Share	MNL 6: Code Share 2	MNL 7: Equip 1	MNL 8: Equip 2	MNL 9: Equip 3
Narrow-body			-0.22 (6.5)	-0.23 (6.6)	-0.23 (6.6)
Commuter			--	--	-1.02 (8.2)
Wide-body (reference)			0	0	0
Presence and Code Share Factors					
Origin presence	0.009 (11)	0.009 (11)	0.009 (9.9)	0.008 (9.8)	0.008 (9.8)
Code share	-2.08 (11)				
Code share – small		-2.82 (4.9)	-2.88 (5.0)	-2.87 (5.0)	-2.87 (5.0)
Code share – large		-1.62 (8.1)	-1.69 (8.5)	-1.68 (8.5)	-1.69 (8.5)
Model Fit Statistics					
LL at zero	-59906.83	-59906.83	-59906.83	-59906.83	-59906.83
LL at convergence	-36729.33	-36708.37	-36510.06	-36527.70	-36528.16
Rho-square zero	0.3869	0.3872	0.3906	0.3903	0.3903
# parameters	23	24	29	27	26
Adjusted rho-square zero	0.3865	0.3868	0.3901	0.3898	0.3898
LRT vs. Model 4	1447,2,<0.001	N/A	N/A	N/A	N/A
LRT vs. Model 5	N/A	41.9,1,<0.001	N/A	N/A	N/A
LRT vs. Model 6	N/A	N/A	397,5,<0.001	N/A	N/A
LRT vs. Model 7	N/A	N/A	N/A	35.3,2,<0.001	N/A
LRT vs. Model 8	N/A	N/A	N/A	N/A	0.92,1,0.34

Note: See Table 7.1 for variable definitions. Carrier constants suppressed for confidentiality reasons. Information provided for Likelihood Ratio Test (LRT) = Likelihood Ratio statistic, degrees of freedom, rejection significance level.

new variables and/or combining categories may cause more dramatic changes in parameter estimates. If large changes in parameter estimates are observed, the analyst should examine the covariance/correlation matrix of parameter estimates (usually included in output or log files).

Even though the likelihood ratio test rejects the null hypothesis that Model 7 = Model 8, Model 8 is the preferred model because it maintains the relationship that propeller aircraft are less preferred than regional jets. However, the parameter estimates for these two equipment types are very similar. Model 9 combines these into a single “commuter” variable. The likelihood ratio test suggests the two models are statistically equivalent, so Model 9 is selected as the preferred model and is used as the basis for exploring market segmentations by departing and returning passengers, direction of travel, and day of week.

Comparison of Outbound and Inbound EW and WE Models

The results discussed to date were for a particular "market segment," specifically for U.S. passengers departing from airports in the Eastern Time Zone and arriving in the Pacific Time Zone (referred to as the EW outbound segment). Table 7.7 shows the results for three other market segments that are differentiated by the direction of travel (EW vs. WE) and whether the passenger is departing from home (outbound) or returning home (inbound). In addition, two "pooled" models are shown so that a formal market segmentation test can be performed. The first "pooled" model is for all passengers traveling from the east coast to west coast (i.e., the data from EW outbound and EW inbound are "pooled" or combined together into a single model). Similarly, a model for all passengers traveling from the west coast to east coast is shown.

By scanning the rows of Table 7.7, one sees that many of the parameter estimates are stable across the four market segments (and the two pooled) models. All but the time of day parameter estimates are similar. This implies that passenger preferences for fare, level of service, aircraft type, code-share, and carrier presence don't vary as a function of the direction of travel (EW vs. WE) or whether they are departing or returning home. However, times of day preferences are influenced by these dimensions, as shown in Figure 7.6.

In the EW direction, departing passengers mostly prefer early morning departures, whereas returning passengers mostly prefer itineraries departing later in the evening. However, in the WE direction, there are three distinct "waves" of preferences. Departing passengers still exhibit the strongest preferences for morning departures, but also show a preference for flights departing late afternoon or overnight red-eye flights. Returning passengers show a similar three-wave pattern, with the strongest preference for flights departing in the late afternoon. The differences in time of day preferences can be due to several factors. First, it is important to remember that when traveling from the east to west coast, passengers gain three hours, so by departing early in the morning, they can still have a productive afternoon on the west coast. In contrast, passengers traveling from the west to the east coast lose three hours, and can't be "productive" in the sense of being able to attend a meeting unless they take a red-eye flight. Second, it is important to note that the model is based on revealed preference data, or current market conditions, and can be influenced by when airlines have scheduled flights. Thus, if airlines did not schedule any red-eye flights, the model would not show any "passenger preference" for red-eye flights. However, given that many flights do not operate at 100 percent load factor (and assuming the price across itineraries is "similar" at the time of purchase), the model based on revealed preference data can be considered as a "fair" representation of passengers' underlying preferences. The analyst may be able to identify potential biases in time of day preferences in the revealed preference data by estimating models for different months (e.g., estimating the model for a high-load factor month of June vs. a low-load factor month of January).

Table 7.7 Comparison of EW and WE segments

	EW Outbound	EW Inbound	EW All pax	WE Outbound	WE Inbound	WE All pax
Carrier Attributes						
Fare ratio	-0.006 (6.3)	-0.008 (10.)	-0.006 (5.9)	-0.006 (8.3)	-0.006 (5.7)	-0.005 (5.0)
Carrier constants	--	--	--	--	--	--
Level of Service						
NS in NS (ref.)	0	0	0	0	0	0
DIR in NS	-2.57 (31)	-2.54 (31)	-2.60 (27)	-2.37 (29)	-2.34 (29)	-2.38 (26)
SC in NS	-4.09 (115)	-3.99 (109)	-4.09 (102)	-4.03 (108)	-3.85 (109)	-3.97 (97)
DC in NS	-9.54 (4.1)	-9.89 (2.6)	-9.70 (2.7)	-9.42 (4.0)	-8.76 (5.1)	-9.07 (3.6)
DIR in DIR (ref.)	0	0	0	0	0	0
SC in DIR	-1.51 (16)	-1.20 (15)	-1.38 (14)	-1.36 (17)	-1.21 (12)	-1.29 (13)
DC in DIR	-6.60 (3.8)	-6.20 (4.9)	-6.44 (3.2)	-6.73 (3.9)	-6.28 (3.0)	-6.52 (2.5)
SC in SC (ref.)	0	0	0	0	0	0
DC in SC	-4.41 (8.0)	-4.25 (9.4)	-4.34 (6.3)	-4.43 (7.9)	-4.11 (7.1)	-4.25 (5.5)
Time of Day						
Sin 2pi	0.046 (1.1)	-0.595 (11)	-0.207 (4.1)	0.134 (4.6)	-0.467 (20)	-0.224 (8.2)
Sin 4pi	-0.291 (6.7)	-0.216 (3.5)	-0.232 (4.3)	-0.240 (8.5)	-0.240 (9.6)	-0.240 (8.1)
Sin 6pi	-0.057 (2.2)	-0.001 (0.0)	-0.010 (0.3)	-0.141 (6.0)	-0.232 (14)	-0.198 (9.4)
Cos 2pi	-0.634 (11)	-0.449 (5.5)	-0.592 (8.2)	-0.174 (6.1)	-0.440 (19)	-0.346 (12)
Cos 4pi	-0.247 (11)	-0.154 (4.3)	-0.247 (8.2)	0.006 (0.2)	0.064 (3.7)	0.001 (0.0)
Cos 6pi	-0.048 (2.8)	-0.065 (3.1)	-0.047 (2.3)	0.126 (5.5)	0.185 (9.2)	0.158 (6.8)
Aircraft Type						
Narrow-body	-0.225 (6.6)	-0.278 (10)	-0.233 (7.0)	-0.305 (12)	-0.247 (7.6)	-0.277 (8.7)
Commuter	-1.023 (8.2)	-1.005 (9.1)	-1.000 (6.9)	-0.986 (8.8)	-0.928 (8.1)	-0.958 (6.9)
Presence and Code Share						
Origin presence	0.008 (9.8)			0.007 (5.4)		
CS depart – small	-2.87 (5.0)			-2.46 (7.5)		
CS depart – large	-1.69 (8.5)			-1.60 (9.2)		
Destination presence		0.010 (7.3)			0.008 (9.7)	
CS return – small		-2.44 (7.6)			-2.86 (5.1)	
CS return – large		-1.50 (9.3)			-1.65 (7.0)	
POS wt. presence			0.006 (4.5)			0.006 (3.7)

Table 7.7 *Concluded*

	EW Outbound	EW Inbound	EW All pax	WE Outbound	WE Inbound	WE All pax
CS – small			-2.55 (7.1)			-2.59 (5.5)
CS – large			-1.81 (7.4)			-1.79 (7.0)
Model Fit Statistics						
LL at zero	-59906.83	-59503.39	-119410.2	-58317.74	-58595.96	-116913.70
LL at convergence	-36528.16	-39624.48	-76761.88	-39224.97	-36660.80	-76120.91
Rho-square zero	0.3903	0.3341	0.3572	0.3274	0.3743	0.3489
# parameters	26	26	26	26	26	26
Adj. rho-square zero	0.3898	0.3338	0.3569	0.3269	0.3739	0.3487

Key: NS = nonstop; DIR = direct; SC = single connection; DC = double connection; CS = code share; POS = point of sale. See Table 7.1 for variable definitions. Carrier constants suppressed for confidentiality reasons.

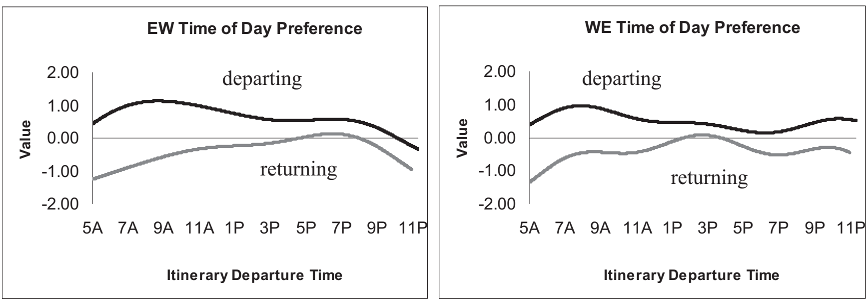


Figure 7.6 Comparison of EW and WE segments

A market segmentation test can be performed to assess whether the improvement in model fit due to differentiating between outbound and inbound passengers is statistically significant. Conceptually, using the EW direction as an example, the null hypothesis is that *all* parameters in the EW outbound model are equal to the parameters in the EW inbound model. Formally:

$$H_0 : \boldsymbol{\beta}_{\text{EW outbound}} = \boldsymbol{\beta}_{\text{EW inbound}}$$

and the decision rule used to evaluate the null hypothesis is given as:

$$\text{Reject } H_0 \text{ if } -2\left[LL_{EWallPax} - LL_{EWout} - LL_{EWin}\right] > \text{critical value from } \chi^2_{NR,\alpha} \text{ distribution}$$

The number of restrictions is equal to the number of parameters, or 26. The null hypothesis is rejected, indicating that models should be differentiated by departing and returning passengers. Formally:

$$-2\times\left[LL\left(EW_{pooled}\right)-\left\{LL\left(EW_{outbound}\right)+LL\left(EW_{inbound}\right)\right\}\right]$$

$$-2\times\left[-76761.11-\left\{-36528.16+\left(-39624.28\right)\right\}\right]>\chi^2_{26,0.05}$$

1218>>39

Similar results apply for the WE-market (470>>39). As a side note, the analyst can estimate "partially pooled" models in which some of the parameters in the EW direction are constrained to be equal (such as fare, level of service, equipment preferences, etc.) and other parameters are allowed to differ (such as time of day, presence, and code-share variables).

Refinement of Time of Day Preferences

Given departing and returning passengers tend to travel on different days of the week, further segmentation by departure day of week may reveal stronger time of day preferences for those days of the week in which business passengers typically fly. One of the concerns, though, that the analyst has to keep in mind, is that further segmentation of the data may lead to model instability due to the smaller number of observations being used to estimate parameters. Tables 7.8 and 7.9 show day of week results for EW outbound and EW inbound itineraries, respectively. Similar to the earlier discussion, many of the parameter estimates (equipment type, carrier presence and code-share factors) are stable across models. In general, level of service variables are stable; however, some estimation problems are starting to appear due to small sample sizes. For example, in the EW inbound model for Saturday, the parameter estimate associated with the "double-connect in non-stop" market returned by the software was unrealistic (-1716) and had no *t*-statistic. An analysis of the data reveals the problem—less than five observations that fall into the EW inbound Saturday segment choose a double-connect when the best level of service is a non-stop. This is one example of "convergence problems" encountered with small sample sizes.

In the EW outbound model, the fare ratio is slightly more negative for itineraries departing on Thursday, Friday, and Saturday and for itineraries returning on Monday and Tuesday, which may be a reflection of larger proportions of price-sensitive leisure passengers departing on these days of the week. Time of day preferences vary across the days of the week, and can be interpreted using the charts in Figure 7.7 that represent the EW airports. Note that the y-axis of all charts has the same scale, to help facilitate comparison across the charts. The charts reveal that EW passengers departing on Monday have strong preferences for early morning flights. This is seen by both the large magnitude in the utility calculated from the six sine and cosine functions, in addition to the lack of an

Table 7.8 EW outbound weekly time of day preferences

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Carrier Attributes							
Fare ratio	-0.003 (1.6)	-0.004 (2.1)	-0.006 (3.4)	-0.009 (5.2)	-0.009 (4.7)	-0.008 (3.8)	-0.005 (2.0)
Carrier constants	--	--	--	--	--	--	--
Level of Service							
NS in NS (ref.)	0	0	0	0	0	0	0
DIR in NS	-2.62 (12)	-2.57 (13)	-2.63 (14)	-2.61 (14)	-2.66 (14)	-2.46 (11)	-2.45 (10)
SC in NS	-4.03 (44)	-4.10 (47)	-4.12 (48)	-4.14 (50)	-4.25 (49)	-4.14 (41)	-3.81 (36)
DC in NS	-9.24 (6.2)	-10.4 (4.2)	-9.40 (6.5)	-9.44 (7.0)	-10.0 (5.7)	-9.37 (6.3)	-9.27 (5.0)
DIR in DIR (ref.)	0	0	0	0	0	0	0
SC in DIR	-1.43 (7.9)	-1.58 (9.5)	-1.54 (9.5)	-1.50 (9.2)	-1.60 (9.4)	-1.67 (8.0)	-1.18 (4.7)
DC in DIR	-6.49 (5.8)	-6.93 (5.5)	-6.64 (6.2)	-6.35 (6.7)	-6.61 (6.0)	-6.98 (4.3)	-6.26 (4.3)
SC in SC (ref.)	0	0	0	0	0	0	0
DC in SC	-4.31 (12)	-4.34 (12)	-4.55 (12)	-4.37 (12)	-4.56 (11)	-4.27 (12)	-4.61 (9.0)
Time of Day							
Sin 2pi	0.20 (1.9)	0.09 (0.8)	0.04 (0.4)	-0.01 (0.1)	0.04 (0.4)	0.08 (0.7)	-0.10 (0.8)
Sin 4pi	-0.27 (2.2)	-0.30 (2.4)	-0.29 (2.5)	-0.34 (3.2)	-0.22 (2.0)	-0.28 (2.1)	-0.26 (1.8)
Sin 6pi	-0.11 (1.3)	-0.10 (1.3)	-0.06 (0.7)	-0.05 (0.7)	0.01 (0.2)	-0.04 (0.5)	-0.05 (0.5)
Cos 2pi	-0.84 (4.1)	-1.00 (4.8)	-0.62 (3.4)	-0.45 (2.8)	-0.38 (2.3)	-0.73 (3.5)	-0.60 (3.1)
Cos 4pi	-0.31 (2.8)	-0.39 (3.3)	-0.22 (2.2)	-0.22 (2.6)	-0.21 (2.4)	-0.22 (1.9)	-0.22 (2.2)
Cos 6pi	-0.03 (0.5)	-0.06 (0.9)	-0.05 (0.9)	-0.07 (1.5)	-0.05 (1.0)	-0.05 (0.8)	-0.03 (0.5)
Aircraft Type (ref=wide-body)							
Narrow-body	-0.26 (2.1)	-0.22 (1.9)	-0.27 (2.4)	-0.24 (2.2)	-0.17 (1.5)	-0.20 (1.5)	-0.19 (1.2)
Commuter	-1.18 (6.6)	-1.09 (6.5)	-1.07 (6.7)	-1.04 (6.6)	-0.86 (5.3)	-0.93 (5.1)	-1.00 (4.6)
Presence and Code Share							
Orig pres	0.013 (6.3)	0.011 (5.9)	0.009 (4.6)	0.005 (2.6)	0.003 (1.4)	0.007 (3.1)	0.015 (5.9)
CS depart – small	-2.97 (6.2)	-3.13 (6.4)	-2.90 (7.0)	-2.62 (7.4)	-2.93 (7.0)	-2.83 (6.1)	-3.02 (4.6)
CS depart – large	-1.67 (5.7)	-1.56 (6.1)	-1.66 (6.5)	-1.73 (7.0)	-1.74 (6.7)	-1.65 (6.0)	-1.90 (5.1)
Model Fit Statistics							
LL at zero	-8364.21	-9102.44	-9709.75	-10404.46	-9573.62	-7115.58	-5636.78
LL at convergence	-5014.07	-5477.11	-5926.59	-6436.90	-5900.49	-4297.17	-3346.40
Rho-square zero	0.4005	0.3983	0.3896	0.3813	0.3837	0.3961	0.4063
# parameters	26	26	26	26	26	26	26
Adj. rho-square zero	0.3974	0.3954	0.3869	0.3788	0.3810	0.3924	0.4017
# Cases	1835	1876	1875	1857	1822	1679	1737

Key: NS = nonstop; DIR = direct; SC = single connection; DC = double connection; CS = code share. See Table 7.1 for variable definitions. Carrier constants suppressed for confidentiality reasons.

Table 7.9 EW inbound weekly time of day preferences

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Carrier Attributes							
Fare ratio	-0.012 (8.6)	-0.011 (6.9)	-0.008 (4.7)	-0.004 (2.8)	-0.001 (0.6)	-0.009 (3.8)	-0.009 (6.1)
Carrier constants	--	--	--	--	--	--	--
Level of Service							
NS in NS (ref.)	0	0	0	0	0	0	0
DIR in NS	-2.53 (15)	-2.51 (13)	-2.51 (12)	-2.57 (12)	-2.65 (11)	-2.28 (7.7)	-2.68 (14)
SC in NS	-4.12 (52)	-4.05 (47)	-4.00 (45)	-3.90 (46)	-3.85 (42)	-3.82 (29)	-4.14 (51)
DC in NS	-9.85 (5.4)	-9.56 (5.2)	-9.67 (4.5)	-9.81 (4.1)	-10.3 (3.0)	-17.16 (-)	-10.1 (4.7)
DIR in DIR (ref.)	0	0	0	0	0	0	0
SC in DIR	-1.33 (8.9)	-1.38 (8.8)	-1.21 (7.5)	-1.09 (6.5)	-0.92 (5.1)	-1.32 (5.7)	-1.14 (6.5)
DC in DIR	-6.40 (6.9)	-6.51 (6.4)	-5.72 (7.6)	-6.18 (6.1)	-6.01 (6.0)	-6.72 (4.1)	-6.25 (6.3)
SC in SC (ref.)	0	0	0	0	0	0	0
DC in SC	-4.25 (13)	-4.28 (12)	-4.16 (12)	-4.17 (12)	-4.38 (11)	-4.18 (9.9)	-4.34 (13)
Time of Day							
Sin 2pi	-0.46 (4.4)	-0.51 (4.2)	-0.72 (5.5)	-0.80 (6.0)	-0.79 (5.2)	-0.54 (3.2)	-0.48 (4.2)
Sin 4pi	-0.14 (1.2)	-0.14 (1.0)	-0.22 (1.4)	-0.32 (2.1)	-0.33 (1.9)	-0.21 (1.1)	-0.26 (1.9)
Sin 6pi	-0.03 (0.3)	0.01 (0.2)	0.03 (0.3)	-0.01 (0.1)	-0.07 (0.7)	0.04 (0.3)	-0.03 (0.4)
Cos 2pi	-0.35 (1.7)	-0.28 (1.2)	-0.50 (2.0)	-0.58 (2.3)	-1.05 (3.3)	-0.20 (0.6)	-0.37 (1.8)
Cos 4pi	-0.15 (1.3)	-0.12 (1.0)	-0.19 (1.4)	-0.24 (1.9)	-0.43 (2.5)	0.18 (1.0)	-0.08 (0.8)
Cos 6pi	-0.12 (2.3)	-0.08 (1.5)	-0.10 (1.6)	-0.07 (1.1)	-0.09 (1.3)	0.09 (1.1)	-0.02 (0.4)
Aircraft Type (ref = wide-body)							
Narrow-body	-0.24 (2.4)	-0.25 (2.2)	-0.30 (2.7)	-0.31 (2.8)	-0.28 (2.3)	-0.13 (0.7)	-0.34 (3.5)
Commuter	-0.93 (6.6)	-0.88 (5.7)	-1.09 (6.8)	-1.10 (7.0)	-1.11 (6.5)	-0.85 (3.6)	-0.99 (6.9)
Presence and Code Share							
Destination pres	0.004 (1.3)	0.008 (2.3)	0.014 (4.2)	0.013 (4.1)	0.015 (4.3)	0.012 (2.3)	0.009 (2.6)
CS return – small	-2.64 (10)	-2.49 (9.0)	-2.28 (8.3)	-2.45 (7.8)	-2.52 (7.5)	-2.30 (5.7)	-2.27 (8.7)
CS return – large	-1.60 (5.6)	-1.36 (4.7)	-1.30 (4.4)	-1.52 (4.9)	-1.61 (4.7)	-2.15 (3.7)	-1.39 (4.8)
Model Fit Statistics							
LL at zero	-10925.93	-8987.51	-8841.02	-9161.93	-7808.19	-3930.67	-9848.13
LL at convergence	-7329.69	-6013.19	-5811.93	-5998.20	-5216.83	-2683.23	-6401.89
Rho-square zero	0.3291	0.3309	0.3426	0.3453	0.3319	0.3174	0.3499
# parameters	26	26	26	26	26	26	26
Adj. rho-square zero	0.3268	0.3280	0.3397	0.3425	0.3285	0.3107	0.3473
# Cases	1835	1876	1875	1857	1822	1679	1737

Key: NS = nonstop; DIR = direct; SC = single connection; DC = double connection; CS = code share. See Table 7.1 for variable definitions. Carrier constants suppressed for confidentiality reasons. Note that the coefficient for a double connection in a nonstop market departing Saturday is unstable (no t-stat and parameter estimate of -17.16) because there are less than five observations for which a double connection is chosen in a nonstop market.

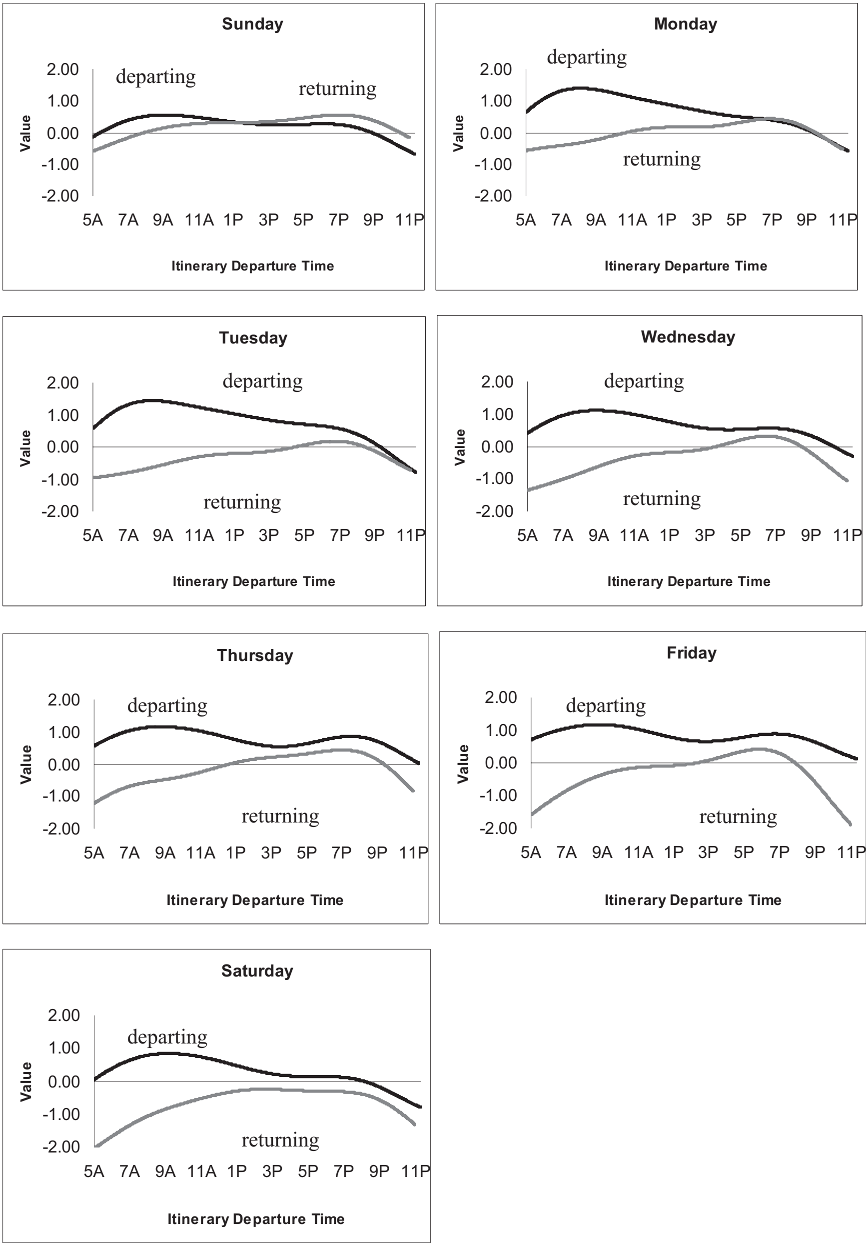


Figure 7.7 Departing and returning time of day preference by day of week

"evening" departure preference seen with some of the other days of the week, such as Thursday and Friday. Passengers departing on Tuesday and Wednesday (which, like those departing on Monday are more likely to be business passengers) also show a preference for early morning flights. Intuitively, this makes sense as passengers departing the east coast of the U.S. during the morning arrive on the west coast in early afternoon, and are still able to have afternoon meetings with clients. Departure time preferences for afternoon flights becomes stronger later in the work week, particularly on Thursday and Friday, which may be a reflection of passenger preferences to depart after work or later in the day for leisure trips. A similar interpretation of time of day preferences is also seen in the returning itineraries. The strongest departure time preferences are seen in the afternoons occurring later in the work week (particularly Thursday and Friday), which may be a reflection of business travelers returning home to the west coast after their meetings have concluded. Among all of the days of the week, Sunday exhibits the weakest time of day preferences.

From a statistical perspective, a market segmentation test can be used to test the null hypothesis that the parameters across the seven days of week are equal. This test rejects the null hypothesis for both the EW outbound and EW inbound data. Formally, the test statistic is given as:

$$-2 \times \left[LL\left(EW_{pooled}\right) - \sum_{DOW_i} LL_i \right] \sim \chi^2_{156,\ 0.05}$$

For the EW outbound data:

$$= -2 \times \left[-36528 - \left\{ -5014 - 5477 - 5927 - 6437 - 5900 - 4297 - 3346 \right\} \right] > \chi^2_{156,\ 0.05}$$

$$260 >> 186$$

Since there are 26 parameters for each model and seven days of week, the number of restrictions is equal to 26×(7-1)=156. Similar results apply for the EW inbound data (338 >> 186).

NL and OGEV Models

From a business perspective, the time of day results by day of week are particularly helpful, as they can help guide decisions on where carriers should schedule flights. However, it is important to note that placing flights during the most popular times of the day and week will not guarantee that the itinerary is profitable. As seen through the MNL models, other factors such as level of service are also very important. Fare, carrier presence, code-share, and equipment type also influence itinerary choice. In addition, the profitability of an itinerary depends on when the other itineraries of the carrier and its competitors are operating in the market. As discussed in Chapter 2, the MNL model imposes the assumption that the

introduction of a new itinerary will draw share proportionately from the itineraries currently operating in that market. However, from an intuitive perspective, one may expect the new itinerary to compete more with other itineraries offered by the carrier in the market and/or other itineraries at the same level of service and/or other itineraries departing around the same time period. More complex models falling into the NL or GNL class can be used to explore whether one or more of these dimensions is important. Table 7.10 summarizes the results of several NL and OGEV models that are discussed in this section.

One of the simplest relaxations of the MNL model is the NL model. Figure 7.8 shows a NL model in which alternatives are grouped into nests according to three time of day categories. This structure reflects the analyst's belief that itineraries departing between 5:00–9:59 compete more with each-other than itineraries departing in the other two time periods: 10AM–3:59 PM and 4:00–11:59PM. The three logsum parameters shown in Figure 7.8 are estimated, and reflect the amount of correlation or substitution among alternatives in the nest. In the EW outbound model, the logsum coefficients are approximately equal, ranging from 0.746 for the first nest to 0.769 for the third nest, which corresponds to correlations of 0.44 to 0.41, respectively. The *t*-statistic associated with each logsum coefficient is greater than 2.0, indicating that the parameter estimate is different from one (the value corresponding to a MNL model) at a significance level of 0.05. The log likelihood also improves, from -36528.16 to -36469.74, and the likelihood ratio test ($-2\times\left[LL_{MNL}-LL_{NL}\right]\sim\chi^2_{3,\,0.05}$) rejects the null hypothesis that the NL and MNL models are equivalent at the 0.05 level since 116.8>>7.8.

Alternative nesting structures may also be possible. Figure 7.9 shows a two-level carrier model in which itineraries are grouped into the same nest if they are operated by the same carrier. Empirical results show that seven of the nine parameter estimates are less than one, indicating that there is a higher degree of substitution among itineraries operated by the same carrier. By definition, carrier 9 represents "all other carriers." Each carrier in this category had less than 5 percent market share across all EW markets. Thus, intuitively, it is not surprising that the logsum coefficient for this nest did not turn out to be less than one. However, because the logsum parameter estimate is greater than one, it is not theoretically valid, and must either be "dropped" from the model (which implies it is constrained to one) or constrained to a value similar to the other nests. Also, in this problem, carrier 6 represents America West, which loosely falls into the category of a low-cost carrier. Passengers who chose this carrier may have been driven more by price than carrier loyalty. Four of the remaining seven parameters have *t*-statistics less than 2.0, indicating that there is not a high degree of substitution among these carriers. Despite the fact that many logsum estimates are not significant at the 0.05 level, the likelihood ratio test of the carrier NL vs. MNL model, which is distributed χ_9^2 , rejects the null hypothesis that the carrier NL and MNL models are equivalent at the 0.05 level since 169>>16.9.

These results are not uncommon for NL models that contain "many" nests. Conceptually, this is due in part to the same sample size issues seen earlier when

Table 7.10 EW outbound NL and OGEV models

	Time	Carrier 1	Carrier 2	Time-Carrier	OGEV
Carrier Attributes					
Fare ratio	-0.005 (6.7)	-0.006 (5.4)	-0.006 (5.4)	-0.005 (5.4)	-0.006 (4.0)
Carrier constants	--	--	--	--	--
Level of Service					
NS in NS (ref.)	0	0	0	0	0
DIR in NS	-2.05 (26)	-2.55 (25)	-2.46 (15)	-2.08 (26)	-2.12 (16)
SC in NS	-3.18 (37)	-3.98 (50)	-3.88 (30)	-3.21 (35)	-3.27 (24)
DC in NS	-7.32 (4.2)	-9.36 (4.1)	-8.96 (2.4)	-7.09 (4.3)	-7.45 (2.4)
DIR in DIR (ref.)	0	0	0	0	0
SC in DIR	-1.18 (14)	-1.34 (13)	-1.43 (8.5)	-1.19 (14)	-1.21 (8.3)
DC in DIR	-5.02 (3.8)	-6.18 (3.9)	-6.18 (2.2)	-4.91 (3.7)	-5.11 (2.2)
SC in SC (ref.)	0	0	0	0	0
DC in SC	-3.42 (7.8)	-4.10 (7.7)	-4.15 (4.5)	-3.36 (7.6)	-3.54 (4.4)
Time of Day					
Sin 2pi	0.086 (2.6)	0.010 (0.2)	0.036 (0.5)	0.085 (2.7)	0.057 (0.9)
Sin 4pi	-0.254 (7.1)	-0.306 (7.0)	-0.278 (3.9)	-0.258 (7.5)	-0.266 (4.2)
Sin 6pi	-0.066 (2.9)	-0.079 (3.1)	-0.058 (1.4)	-0.066 (3.1)	-0.039 (1.0)
Cos 2pi	-0.525 (11)	-0.678 (11)	-0.603 (6.3)	-0.508 (11)	-0.548 (6.5)
Cos 4pi	-0.220 (11)	-0.264 (10)	-0.233 (6.0)	-0.229 (12)	-0.210 (6.2)
Cos 6pi	-0.029 (2.1)	-0.051 (3.1)	-0.046 (1.7)	-0.023 (1.7)	-0.013 (0.5)
Aircraft Type (ref=wide-body)					
Narrow-body	-0.150 (5.3)	-0.213 (3.8)	-0.213 (3.8)	-0.133 (4.7)	-0.153 (3.0)
Commuter	-0.773 (7.8)	-0.921 (7.8)	-0.966 (4.6)	-0.749 (7.5)	-0.788 (4.5)
Presence and Code Share					
Origin presence	0.006 (9.3)	0.008 (8.8)	0.010 (6.1)	0.010 (12)	0.006 (5.3)
CS depart – small	-2.18 (4.9)	-3.12 (5.8)	-2.76 (2.9)	-2.25 (5.2)	-2.20 (2.8)
CS depart – large	-1.26 (8.2)	-1.59 (7.8)	-1.54 (4.5)	-1.17 (7.7)	-1.28 (4.7)
NL Logsums					
TOD Nest1	0.746 (6.0)			0.857 (3.3)	
TOD Nest 2	0.752 (5.3)			0.900 (2.1)	

Table 7.10 *Concluded*

	Time	Carrier 1	Carrier 2	Time-Carrier	OGEV
TOD Nest 3	0.769 (5.4)			0.886 (2.6)	
Carrier – Nest 1		0.850 (3.1)	0.925 (1.5)	0.678 (6.5)	
Carrier – Nest 2		0.783 (3.7)	0.925 (1.5)	0.678 (6.5)	
Carrier – Nest 3		0.993 (0.1)	0.925 (1.5)	0.678 (6.5)	
Carrier – Nest 4		0.880 (2.1)	0.925 (1.5)	0.678 (6.5)	
Carrier – Nest 5		0.948 (1.3)	0.925 (1.5)	0.678 (6.5)	
Carrier – Nest 6		1.241 (8.6)	0.925 (1.5)	0.678 (6.5)	
Carrier – Nest 7		0.989 (0.2)	0.925 (1.5)	0.678 (6.5)	
Carrier – Nest 8		0.913 (1.5)	0.925 (1.5)	0.678 (6.5)	
Carrier – Nest 9		1.809 (23)	0.925 (1.5)	0.678 (6.5)	
OGEV					
α_0					0.41 (1.6)
Logsum for TOD					0.758 (3.7)
Model Statistics					
LL at zero	-59906.83	-59906.83	-59906.83	-59906.83	-59906.83
LL at convergence	-36469.74	-36359.63	-36519.40	-36373.43	-36461.70
Rho-square zero	0.3912	0.3931	0.3904	0.3928	0.3916
# parameters	29	35	27	30	28
Adj. rho-square zero	0.3907	0.3925	0.3900	0.3923	0.3914

Key: NS = nonstop; DIR = direct; SC = single connection; DC = double connection; CS = code share. See Table 7.1 for variable definitions. Carrier constants suppressed for confidentiality reasons. T-stats for logsum and allocation parameters reported against 1.

the dataset was divided into seven separate market segments, one for each day of the week. In this case, nine logsum parameters are being estimated, so if the frequency of alternatives chosen in the nest is low, it can be difficult to obtain significant and/or stable parameter estimates. One of the ways that is commonly used to correct for this instability is to constrain logsum parameters to be the same across nests. This result is shown as the “Carrier 2” model in Table 7.10. In this case, the logsum coefficient is 0.925, which is close to one (or a MNL model). Comparing the time NL and Carrier 2 NL models, one concludes that time of

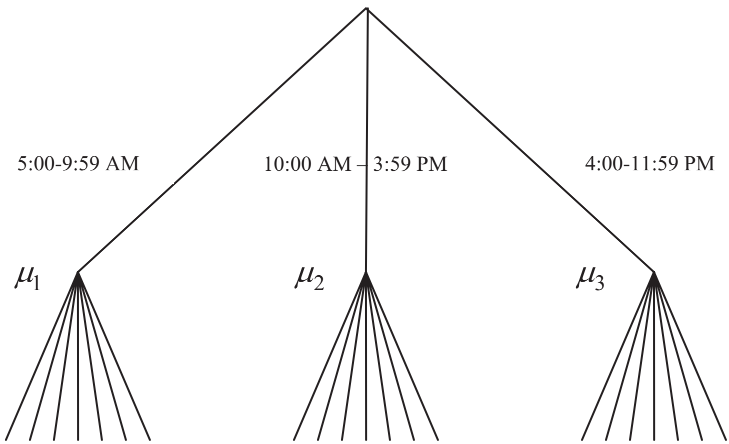


Figure 7.8 Two-level NL time model structure

day substitution is stronger than carrier substitution. As a final note, based on the results of the Carrier 1 NL model, the analyst may elect to constrain the logsums to be equal for all carriers *except* six and nine, and impose the MNL restriction on these latter two to reflect the analyst's belief that low cost carriers and carriers with small market shares do not exhibit a strong brand presence / intra-competition level. The non-nested hypothesis test can be used to determine which model fits the data better.

The time of day and carrier structures are "two level" models in the sense that alternatives are grouped into nests according to only one dimension. Figure 7.10 shows a three-level NL model that groups nests by time of day (at the upper level) and carrier (at the lower level). Given this structure results in 27 carrier logsum parameters, these are constrained to be equal to each other.

Empirical results indicate that the logsum coefficients associated with time of day (0.857, 0.900, and 0.886) are statistically different than one at the 0.05 level. Most important, the carrier logsum coefficient of 0.678 is less than the time of day logsum coefficients. This is a theoretical requirement of the model, as carrier logsums greater than 0.857 would imply a negative variance in the model, which is not possible. The three-level NL model indicates that itineraries that share the same operating carrier and departure time category compete most with each other, followed by those itineraries in the same departure time period. Conceptually, the level of competition between two itineraries can be thought of in terms of which "nodes" connect them. If the path connecting two itineraries can only be drawn by going through the root node at the top of the tree, the MNL proportional substitution property holds. Alternatives that share the same, lower-level nest (i.e., can be connected by only using a carrier node) exhibit the greatest substitution. Finally, alternatives that share the same upper-level nest but different lower-level nests (i.e., are

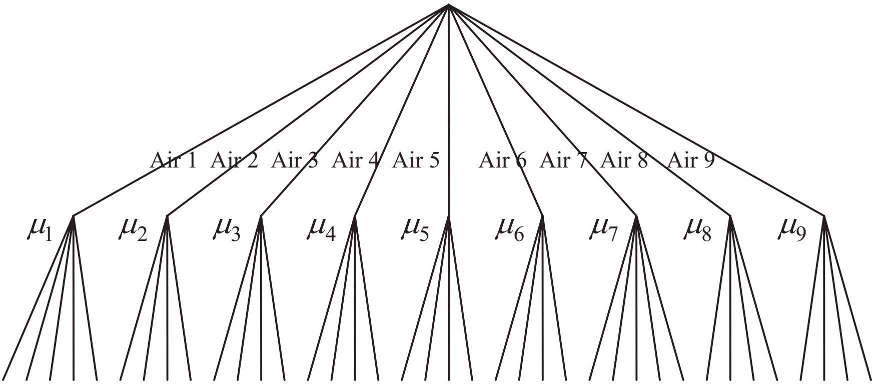


Figure 7.9 Two-level carrier model structure

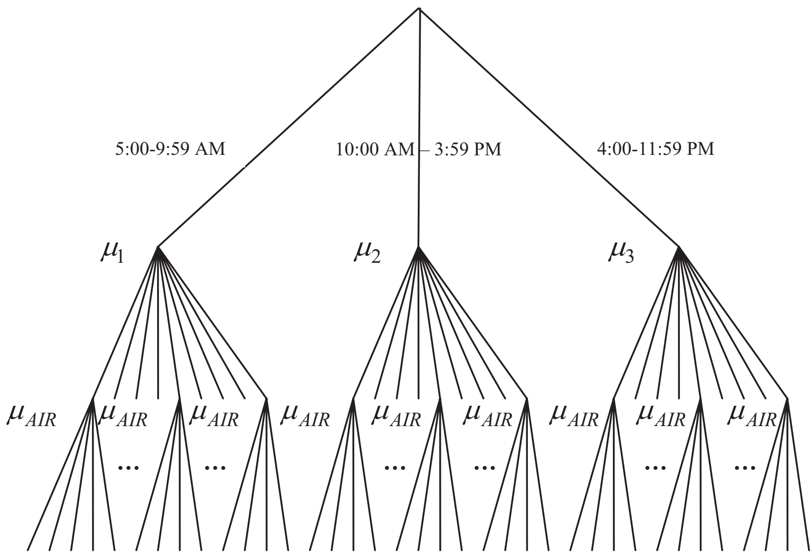


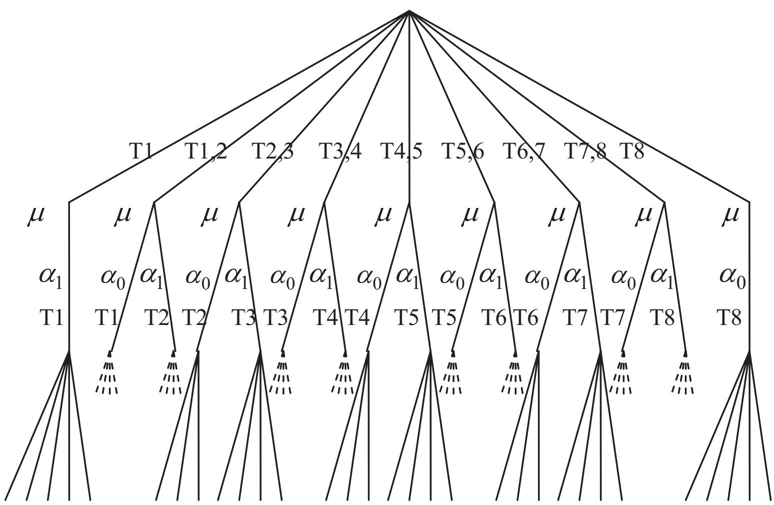
Figure 7.10 Three-level time-carrier model structure

connected by passing through two carrier nodes and one time of day node) fall between the other two cases.

One of the problems with the models discussed thus far is that the three time of day nests are “arbitrary” in the sense that different breakpoints for time of day could have been used (and may lead to different results). In addition, the use of breakpoints implies that itineraries departing at 9:50 A.M. and 9:58 A.M. exhibit increased competition between each other, but that itineraries departing at 9:58 A.M. and 10:02 A.M. exhibit proportional substitution (i.e., they do not share the same nest because a breakpoint of 10 AM was used). In addition, within a nest,

itineraries compete equally with each other. However, intuitively, the 8:30 A.M. departures are expected to compete more with the 8:00 and 9:00 departures than the 7:30 and 9:30 departures. The ordered generalized extreme value (OGEV) model can be used to partially overcome these problems. The OGEV model is a special case of the GNL model. Specifically, alternatives are allocated to multiple nests according to their proximity to each other along the time of day dimension. This structure is shown in Figure 7.11.

The OGEV model is similar to the time NL model. First, the logsum coefficients in the OGEV model were constrained to be equal to each other. The OGEV logsum coefficient (0.758) is approximately equal to the logsums of the Time NL model with three nests (0.746 to 0.769). Consistent with expectation, the allocation parameter is approximately equal to 0.5, that is, one would not expect an itinerary departing at 10 A.M. to draw disproportionately higher share from the 8 A.M. flight than the noon flight (this would occur if the allocation parameter, α_0, were close to 1). The non-nested hypothesis test of the Time NL versus OGEV model rejects the null hypothesis that the two models are equal at a significance << 0.001. Thus, based on the fact that the OGEV model provides a better behavioral representation and fits the data better, it is the preferred model.



Note: T1 = 5-6:59 AM; T2 = 7-8:59 AM; T3 = 9-10:59 AM; T4 = 11-12:59PM; T5 = 1-2:59PM; T6 = 3-4:59PM; T7 = 5-6:59PM; T8 = 7-11:59PM

Figure 7.11 OGEV model structure

Bringing it All Together… Which Model Should We Use?

This section illustrated the discrete choice modeling process for airline itinerary choices. The modeling process began with descriptive statistics and estimation of MNL models that evolved from a simple specification to a more complex one. The selection of a "preferred" MNL model was guided by formal statistical tests (*t*-tests, likelihood ratio test, non-nested hypothesis test), business requirements (e.g., the desire to capture differences in code-share flights by whether the carrier has a large or small market presence), and analyst intuition (e.g., it is expected that regional jets will be preferred to propeller aircraft). The preferred model specification was applied to various market segments, differentiated by direction of travel (EW vs. WE), departing vs. returning passengers, and/or day of week. It was only after this stage that more advanced discrete choice models, specifically NL and OGEV models, were estimated.

Results of the NL and OGEV models indicate that they fit the data better than the MNL model. In general, parameter estimates associated with the independent variables were similar across all models (MNL, NL, and OGEV). Although parameter estimates can change when a more advanced discrete choice model is used, in most applications they are relatively stable (with the exception of alternative-specific constants). Thus, the key benefit of using the NL and OGEV models is the ability to incorporate more realistic substitution patterns among alternatives.

From a practical implementation perspective, however, it is important to note that although the NL and OGEV models incorporate more realistic substitution patterns, their probability expressions are more complex and require additional information of which alternatives belong to which nest. Thus, there is a trade-off between using the "simple" MNL model and more complex NL or OGEV models. Based on the experience of Gregory Coldren in implementing MNL models for major U.S. airlines (Coldren Koppelman Kasturirangan and Mukherjee 2003) the MNL model has been seen to offer dramatic improvements in model fit over QSI-based models; forecasting gains from using the more complex models are likely to be less (but have not yet been examined). Further, it is our opinion that using a MNL model and combining data from multiple sources (namely booking data and on-line shopping data) will lead to larger improvements in forecasting accuracy than developing more complex discrete choice models. This is because the on-line shopping data contain information on actual (and "disaggregate") prices and itineraries shown to the customers, not quarterly fare information.

Conclusions

This chapter focused on describing two major types of market share models found in scheduling systems: those based on the QSI methodology and those based on discrete choice methods.

Based on interviews with industry experts, we learned that many airlines currently using logit-based methodologies are contemplating reintroducing QSI methodologies due to the perceived complexity of logit models and difficulty in maintaining parameter estimates. However, in our opinion, we believe that the fundamental problems currently being observed are not due to the use of a logit model, but rather over-parameterized utility functions, particularly those that use schedule delay functions. One of the primary differences between the published MNL model of a major U.S. carrier (which was clearly seen to dominate their QSI model) and the logit models used in practice relates to the number of variables (and estimated β coefficients). In the published MNL model, each regional entity has 36 parameter estimates in addition to estimates associated with each airline carrier. Further, 18 of these parameters, which are associated with dummy variables for time of day preferences, can be further reduced via the use of the continuous sine and cosine functions described in this chapter. This is in comparison with alternative logit models reported to have hundreds, *if not thousands*, of parameter estimates. However, a simple, yet well-specified MNL utility function can lead to superior predictive performance over a QSI model. Complexity should not be driven by the number of variables included in the model, but rather by the desire or need to obtain more accurate substitution patterns than those imposed by the MNL. If desired, more flexible patterns can be incorporated via the use of more advanced models, such as the NL and OGEV models presented in this chapter, the more general Weighted NL and Nested Weighted NL models discussed in Chapter 4 that belong to the NetGEV family (Chapter 5), or the mixed logit models (Chapter 6).

Summary of Main Concepts

This chapter presented the modeling process that is used to develop a well-specified utility function and relax the IIA property associated with the MNL model. The most important concepts covered in this chapter include the following:

- Several statistics are used in discrete choice modeling to help guide the selection of a preferred model. The *t*-statistic is used to test a null hypothesis related to a single parameter estimate. Non-nested hypothesis tests are used to compare two models. The likelihood ratio test is used when one model can be written as a restricted version of a different model. Non-nested hypothesis tests are used when one model cannot be written as a restricted version of a second model.
- In discrete choice models, rho-squares ρ^2 and adjusted rho-squares $\bar{\rho}^2$ provide information about the goodness of fit of a model (and play an analogous role to R^2 and adjusted R^2 measures used in linear regression).
- Two common ρ^2 reference models include an "equally likely model" and a "market shares model." In an equally likely model, each alternative in the choice set is assumed to have an equal probability of being chosen.

A market share reference model is a model that contains a full set of identified alternative-specific constants. The constants-only model assumes each alternative has a probability of being selected that corresponds to the sampling shares.

- Rho-squares are a descriptive—and subjective—measure that will be sensitive to the frequency of chosen alternatives in the sample.
- The selection of a "preferred" MNL model specification is guided by formal statistical tests, business requirements, and analyst intuition.
- The majority of an analyst's modeling time is spent refining the utility function. More advanced models (such as the NL or OGEV models) are estimated after a preferred utility function has been selected. Although parameter estimates can change when a more advanced discrete choice model is used, in most applications they are relatively stable.
- This chapter highlighted several common issues analysts are likely to encounter when estimating models. Small sample sizes can lead to models that fail to converge, fail to return *t*-statistics, and/or return parameter estimates that are very large in absolute magnitude. When estimating advanced models, it is common to constrain logsum coefficients and/or allocation parameters to be equal to each other, particularly when the number of nests is large and/or some nests have alternatives that are infrequently chosen.
- This chapter also highlighted several common estimation "tricks." For example, the dependent variable representing the number of itinerary bookings was re-scaled to fall in the [0,1] interval, which decreased estimation time by a factor of four. It is also common to estimate logsum parameters without constraining them to fall in the (0,1] range and reject models that have logsum coefficients outside this range. A trial-and-error method for non-linear-in-parameters function was also discussed. All of these estimation "tricks" point to a research need for more robust nonlinear algorithms to solve for the parameters of discrete choice models.
- The term "nested logit model" is sometimes used incorrectly in the context of airline itinerary choice models to refer to a MNL model that has a schedule delay function.
- The authors believe that the fundamental problems currently being observed in practice related to logit-based itinerary choice models are due to over-parameterized utility functions that incorporate schedule delay functions; these models often have thousands of parameters.

This page has been left blank intentionally

Chapter 8
Conclusions and Directions for Future Research

Introduction

Chapter 1 described how the different backgrounds, perspectives, and operational requirements between the aviation and urban travel demand modeling areas resulted in different demand forecasting practices. It was not until the early 2000's that operations research (OR) analysts and aviation practitioners began to openly express interest in using discrete choice models to represent individual consumers' choices. This book has focused on those theoretical developments and applications that will be important foundations for the next decade of demand modeling in the airline industry. This chapter will provide perspectives on additional research opportunities emerging in the airline industry that can leverage the benefits of discrete choice and other behavioral models based on disaggregate data. These research opportunities span revenue management, pricing and new product development.[1]

Revenue Management Applications

To date, the majority of aviation applications using discrete choice models with revealed preference data have focused on either applications in which it is relatively easy to generate the choice set (e.g., itinerary choice problems) or applications in which it is relatively easy to replace a forecasting component (e.g., a no show model) that is part of a much larger decision-support system (e.g., a revenue management system). Researchers have started investigating benefits associated with developing the next generation of choice-based revenue management (RM) systems. Choice-based RM systems seek to replace demand forecasting models based on probability and time-series methods with demand forecasting models based on discrete choice methods. This is a challenging problem, and one that requires substantial, multi-million dollar investments, due to the need to collect and maintain information on not only the product purchased by an individual, but also the menu of choices the individual viewed prior to purchase.

1 This chapter draws heavily on an article that appeared in the *Journal of Revenue and Pricing Management*, and has been reproduced with permission of Palgrave Macmillan (Garrow 2009).

Looking ahead, there are several challenges that will need to be addressed in order to operationalize choice-based RM methods. The methods used to represent individuals' willingness to pay will have to be revisited. Specifically, two distinct behavioral components will need to be modeled: an individual's willingness to pay to travel by air (or market size) and an individual's willingness to pay for a specific itinerary (or market share). As currently structured, many choice-based RM formulations construct a choice set that includes a "no purchase" alternative, and assign a utility value of zero to this alternative. However, from a behavioral perspective, this is not realistic, as one firm's no-purchase alternative is likely a different firm's customer. Incorporating measures of competitor prices and/or developing separate models for market size and market share will likely be required in order to effectively operationalize choice-based RM models. In addition, it is important to remember that parameter estimates for discrete choice models are (typically) obtained via maximum likelihood methods. These methods often result in under-forecasting of alternatives that are infrequently chosen (such as high yield products). This suggests more sophisticated estimation techniques and/or posterior processing techniques may need to employed to ensure the model is not under-forecasting these valuable customers, thereby contributing to another realization of the spiral-down phenomena (e.g., see Boyd and Kallesen 2004).

Although developing new RM methods to account for increased fare competition is an important research direction, the Internet has also resulted in a variety of new needs to enhance traditional RM modules. For example, calibrating models by customer segments to distinguish between time-sensitive and price-sensitive customers is a potentially high yield research topic; that is, increased fare competition and the elimination of many fare product restrictions has resulted in a blurring of these segments within booking classes. New segmentation schemes based on "round trip" information available at the time of booking, such as those investigated by Carrier (2008), may be one way to develop more accurate demand forecasts that distinguish between price-sensitive and time-sensitive customers. Indeed, despite the fact that many airlines have transitioned to one-way pricing, jointly considering outbound and inbound itinerary information (captured at the time of booking) will likely greatly enhance forecasting accuracy. Over the last five years, work spanning air travelers' no show, cancellation, and itinerary choices have consistently shown substantial differences in how individuals value the attributes of outbound and inbound itineraries (e.g., see Garrow and Koppelman 2004a, 2004b; Koppelman, Coldren and Parker 2008; Iliescu, Garrow and Parker 2008). From individuals' time-of-day preferences to rescheduling behavior, empirical evidence shows that customers' trip scheduling requirements are more rigid on the outbound portion (thereby implying a higher willingness to pay for a preferred outbound itinerary versus inbound itinerary).

Online travel agencies are particularly well positioned to make advancements in recapture rate modeling. This is due to the fact that their menus contain itineraries representing different carriers, times of day, prices, and levels of service. In this context, discrete choice models provide a rich set of insights related to

cross-elasticity and substitution effects. Further, if one is willing to assume that the independence of irrelevant alternative (IIA) property holds, recapture rates (as reflected in the redistribution of choice probabilities when an alternative is no longer available) become very straightforward to calculate. Although the IIA property is likely violated for itinerary choice applications, this methodology nonetheless represents a substantial improvement over recapture rate methods currently used in practice. Extensions to more advanced logit specifications, as well as consideration of outbound and inbound differences in recapture rates, are all valuable research extensions to work that has been done in this area (e.g., see Ratliff, Venkateshwara, Narayan and Yellepeddi 2008).

Pricing Applications

Perhaps one of the most controversial questions related to the Internet is whether models should be developed that predict competitors' prices. From an industry perspective, integrating detailed price forecasts into a traditional RM framework is fraught with danger, i.e., increased knowledge of competitors' pricing behavior may result in increased competition and smaller profits. On the other hand, in certain applications, knowing individuals' willingness to pay for product attributes, particularly with respect to competitors' offerings, will be quite valuable (as is the case with recapture rate models or flight scheduling models).

There is growing interest in modeling competition and the interaction between optimal firm pricing policies and customer decisions. Interest in this area spans multiple areas, including the OR community (which has shown a particular interest in understanding the role of strategic customers), classic economics (which has tended to focus on analyzing market impacts of mergers and code-shares), and behavioral economics. Competitive response models that jointly incorporate customer search and purchase behaviors are a potentially fruitful area of research. That is, the ability to track individuals and observe—unobtrusively—how they search for information on the Internet prior to purchase provides a unique research opportunity. Further, by framing the classic game theory problem as one of search and purchase and extending dynamic discrete choice models to incorporate time-varying prices, customer behavior assumptions are not trivial and in some cases can lead to pricing forecasts that are opposite those predicted by classic models (Castillo, Garrow and Lee 2008). Models that jointly consider search and purchase may have implications to a wide range of applications, including merger and acquisition studies.[2] Insights gained from jointly investigating customer search and purchase behaviors may also lead to new product and/or menu designs.

---

2 Price dispersion is one measure that is frequently examined when analyzing the impacts of mergers and acquisitions. Theoretically, price dispersion can occur when there are non-zero search costs.

New Product Development

In many ways, the Internet has been both a blessing and a curse for carriers. On one hand, carriers have benefited from lower distribution costs and the ability to interact directly with customers (versus relying on an intermediary travel agency). On the other hand, the Internet has not only increased the transparency of prices for customers, but for competitors as well. Monitoring competitive prices and seat availability (a measure of demand on competitors' flights) is becoming more common and viable at a large scale. The net result is a highly competitive market in which the ability to segment customers and price discriminate is becoming more difficult and price changes are quickly matched by competitors. In this environment, one may question whether it is possible for a carrier to leverage the strengths of the Internet—and specifically the ability to interact directly with its consumers—to customize prices for individual consumers in ways that do not trigger price responses by the competition. Conceptually, the fundamental question of interest is to determine whether it is possible to stimulate new leisure demand by designing a product for highly time-flexible travelers that is sold via the Internet.

In May of 2003, David Post launched a web-based Interactive Pricing Response (IPR) system with Freedom Air International, a former low cost subsidiary of Air New Zealand, to explore these questions. This IPR system enables customers to generate prices individually based on their level of time-flexibility. The IPR system is designed to tap into a predominately unserved market of highly time-flexible people that would fly if they were able to offset some of their time-flexibility for a larger discount than is presently offered by the airlines. This allows airlines both the ability to generate incremental revenues, as well as the ability to better match supply and demand. Most important, competitors have no set price target on which to compete because prices are customized to individual consumers and are dynamically generated based on the airline's current demands (Post, Mang and Spann 2007). In this context, Freedom Air was able to generate incremental revenues by effectively making its discount levels opaque to competitors, despite the fact the IPR system operated in an online distribution channel. IPR also enabled Freedom Air to "re-segment" the market and price discriminate based on the travelers' degree of time-flexibility. As the airline industry becomes even more competitive, and traditional product characteristics such as Saturday night stay, advance purchase, and other restrictions become obsolete, finding more innovative ways to price discriminate becomes even more crucial, particularly if fuel costs continue to rise.

Although the original vision for the IPR system was to provide a mechanism by which legacy airlines could better compete with low cost carriers without engaging in pricing wars, it is interesting to note that smaller, low cost airlines outside the U.S. have been the early adopters. Looking ahead, it will be interesting to see if this product, or others that leverage the unique strengths of the Internet,

are able to penetrate the market and, if so, which areas of the world will be most receptive to its implementation.

Regulatory and Other Factors

Perhaps the most significant factors that will influence future business needs and research opportunities over the next decade are new regulations and domestic and foreign policies. Dramatic increases in fuel costs, the discovery of an alternative source of energy, and slowing of airport growth in urban areas due to carbon emissions caps would have dramatic impacts on airline costs and market competition.

One thing that is certain, however, is that the airline industry is very dynamic. A decade from now, it is almost certain that researchers will be challenged to solve revenue management, pricing, scheduling, marketing, and operations problems that need to incorporate factors that were not anticipated today. In this context, developing demand forecasting models that incorporate important elements of customer behavior while simultaneously enabling these decision-support systems to identify and respond to unanticipated market changes will be important. Indeed, investigating methods to forecast airline demand and integrating new behavioral findings into larger decision-support systems are fruitful and exciting areas of research, and ones that can only be strengthened through collaborations between OR and discrete choice modeling analysts.

This page has been left blank intentionally

References

Aaron, K. 2007. SITA survey: Choice is key to encouraging self-service. *Air Transport World*, 76–80 [Online]. Available at: www.accessmylibrary.com/coms2/summary_0286-33306606_ITM [accessed: 9 May 2009].

Abbe, E., Bierlaire, M. and Toledo, T. 2007. Normalization and correlation of cross-nested logit models. *Transportation Research Part B*, 41(7), 795–808.

Adler, T., Falzarano, C.S. and Spitz, G. 2005. Modeling service trade-offs in air itinerary choices. *Transportation Research Record,* 1915, 20–6.

Airline Deregulation Act. 1978. In *Public Law No. 95–504, 92 Stat. 1750*.

Airlines Reporting Corporation. 2009. *Company Information* [Online]. Available at: http://www.arccorp.com/index.jsp [accessed: 2 June 2009].

Akar, G., Clifton, K.J. and Doherty, S.T. 2009. *How Do Travel Attributes Affect the Planning Time Horizon of Activities?* Paper to the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

Algers, S. and Beser, M. 2001. Modelling choice of flight and booking class—A study using stated preference and revealed preference data. *International Journal of Services Technology and Management*, 2(1/2), 28–45.

Allenby, G. and Ginter, J. 1995. The effects of in-store display and feature advertising on consideration sets. *International Journal of Research in Marketing*, 12, 67–80.

ALOGIT Software & Analysis Ltd. 2008. *ALOGIT Version 4.2* [Online]. Available at: http://www.alogit.com [accessed: 21 September 2008].

Aptech Systems Inc. 2008. *Gauss Version 9.0* [Online]. Available at: http://www.aptech.com [accessed: 21 September 2008].

Amadeus. 2009. *Company Information* [Online]. Available at: http://www.amadeus.com/amadeus/amadeus.html [accessed: 2 June 2009].

Ashiabor, S., Baik, H. and Trani, A. 2007. Logit models for forecasting nationwide intercity travel demand in the United States. *Transportation Research Record*, 2007, 1–12.

Bastin, F., Cirillo, C. and Toint, P.L. 2003. *Numerical Experiments with AMLET, a New Monte-Carlo Algorithm for Estimating Mixed Logit Models.* To the 10th International Conference on Travel Behaviour Research, Lucerne, Switzerland.

Bastin, F., Cirillo, C. and Toint, P.L. 2009. Estimating mixed logit with non-parametric random variables. *Working Paper*. Available at: http://www.cerfacs.fr/algor/reports/2006/TR_PA_06_87.ps.gz [accessed 1 April 2009].

Beckor, S., Ben-Akiva, M. and Ramming, M. 2002. Adaptation of logit kernel to route choice situation. *Transportation Research Record*, 1805, 78–85.

Bellemans, T., van Bladel, K., Janssens, D., Wets, G. and Timmermans, H. 2009. *Measuring and Estimating Suppressed Travel Using Enhanced Activity-Travel Diaries*. Paper to the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

Ben-Akiva, M. 1974. Note on the specification of a logit model with utility functions that include attributes of competing alternatives. *Working Paper*, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Ben-Akiva, M., Bolduc, D. and Bradley, M. 1993. Estimation of travel choice models with randomly distributed values of time. *Transportation Research Record,* 1413, 88–97.

Ben-Akiva, M., Bolduc, D. and Walker, J. 2001. Specification, identification, and estimation of the logit kernel (or continuous mixed logit) model. *Working Paper*, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Ben-Akiva, M. and Francois, B. 1983. μ-homogeneous Generalized Extreme Value model. *Working Paper*, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Ben-Akiva, M. and Lerman, S.R. 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. Cambridge: The MIT Press.

Ben-Yosef, E. 2005. The Evolution of the US Airline Industry: Theory, Strategy, and Policy. Dordrecht: Springer.

Besbes, O. and Zeevi, A. 2006. Implications of choice model parameter estimation in a competitive environment. *Working Paper*, Stern School of Business, New York University.

Bhat, C.R. 1995. A Heteroscedastic Extreme Value model of intercity mode choice. *Transportation Research Part B*, 29(6), 471–83.

Bhat C.R. 1997. A nested logit model with covariance heterogeneity. *Transportation Research Part B*, 31(1), 11–21.

Bhat, C. 1998a. Accommodating flexible substitution patterns in multidimensional choice modeling: Formulation and application to travel model and departure time choice. *Transportation Research Part B*, 32(7), 455–66.

Bhat, C. 1998b. Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. *Transportation Research Part A*, 32(7), 495–507.

Bhat, C.R. 2000a. Flexible structures for discrete choice analysis, in *Handbook of Transport Modeling*, edited by D.A. Hensher and K.J. Button. Amsterdam: Pergamon, 71–90.

Bhat, C. 2000b. Incorporating observed and unobserved heterogeneity in urban work model choice modeling. *Transportation Science*, 34, 228–38.

Bhat, C. 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B*, 35(7), 677–93.

Bhat, C. 2003a. *Econometric Choice Formulations: Alternative Model Structures, Estimation Techniques, and Emerging Directions*. Paper to the Tenth International Conference on Travel Behaviour Research, Lucerne, Switzerland.

Bhat, C. 2003b. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B*, 37(9), 837–55.

Bhat, C.R. and Castelar, S. 2002. A unified mixed logit framework for modeling revealed and stated preferences: Formulation and application to congestion pricing analysis in the San Francisco Bay Area. *Transportation Research Part B,* 36(7), 593–616.

Bhat, C.R., Eluru, N. and Copperman, R.B. 2008. Flexible model structures for discrete choice analysis, in *Handbook of Transport Modeling* edited by D.A. Hensher and K.J. Button. 2nd Edition. Amsterdam: Elsevier, 75–104.

Bhat, C. and Gossen, R. 2004. A mixed multinomial logit model analysis of weekend recreational episode type choice. *Transportation Research Part B*, 38(9), 767–87.

Bierlaire, M. 2003. *BIOGEME: A Free Package for the Estimation of Discrete Choice Models.* Paper to the Third Swiss Transportation Research Conference, Ascona, Switzerland.

Bierlaire, M. 2008. *An Introduction to BIOGEME Version 1.7* [Online]. Available at: http://www.biogeme.epfl.ch [accessed: 21 September 2008].

Bierlaire, M., Bolduc, D. and McFadden, D. 2008. The estimation of Generalized Extreme Value models from choice-based samples. *Transportation Research Part B,* 42(4), 381–94.

Bodea, T.D., Ferguson, M.E. and Garrow, L.A. 2009. Choice-based revenue management: Data from a major hotel chain. *Manufacturing and Service Operations Management* 11(2): 356–61.

Bodea, T.D. and Garrow, L.A. 2006. *The Importance of Synthetic Datasets in Empirical Testing: Comparison of NL and MNL with Error Components Models*. Paper to the *European Transport Conference*, Strasbourg, France.

Bolduc, D., Fortin, B. and Fournier, M. 1996. The effect of incentive policies on the practice location of doctors: A multinomial probit analysis. *Journal of Labor Economics*, 14(4), 703–32.

Bowman, J. 2004. *Heteroscedastic Nested Logit Kernel (or Mixed Logit) Models for Large Multidimensional Choice Problems: Identification and Estimation.* Paper to the 83rd Annual Meeting of the Transportation Research Board, Washington, D.C.

Boyd, J. and Mellman, J. 1980. The effect of fuel economy standards on the U.S. automotive market: A hedonic demand analysis. *Transportation Research Part A*, 14, 367–78.

Boyd, A. E. and Kallesen, R. 2004. The science of revenue management when passengers purchase the lowest available fare. *Journal of Revenue and Pricing Management,* 3(2), 171–77.

Braaten, E. and Weller, G. 1979. An improved low-discrepancy sequence for multi-dimensional quasi-Monte Carlo integration. *Journal of Computational Physics*, 33, 249–58.

Bresnahan, T.F., Stern, S. and Trajtenberg, M. 1997. Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s. *RAND Journal of Economics*, 28, S17-S44.

Bront, J.J.M., Mendez-Diaz, I. and Vulcano, G. 2007. A column generation algorithm for choice-based network revenue management. *Working Paper*, Stern School of Business, New York University.

Brownstone, D. 2001. Discrete choice modeling for transportation, in *Travel Behavior Research: The Leading Edge*, edited by D.A. Hensher. Oxford, UK: Elsevier, 97–124.

Brownstone, D., Bunch, D. and Train, K. 2000. Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B,* 34, 315–38.

Brownstone, D. and Small, K. 2003. *Valuing Time and Reliability: Assessing the Evidence from Road Pricing Demonstrations.* Paper to the Tenth International Conference on Travel Behaviour Research, Lucerne, Switzerland.

Brownstone, D. and Train, K. 1999. Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics*, 89, 109–29.

Bunch, D.A. 1991. Estimability in the multinomial probit model. *Transportation Research Part B*, 25: 1–12.

Bureau of Transportation Statistics, U.S. Department of Transportation. 1995. *American Travel Survey: An Overview of the Survey Design and Methodology.*

Bureau of Transportation Statistics, U.S. Department of Transportation. 2009. O*rigin and Destination Data Bank* [Online]. Available at: http://www.bts.gov [accessed: 8 May 2009].

Calfee, J., Winston, C. and Stempski, R. 2001. Econometric issues in estimating consumer preferences from stated preference data: A case study of the value of automobile travel time. *Review of Economics and Statistics,* 83(4), 699–707.

Cardell, N.S. 1997. Variance components structures for the Extreme Value and Logistic distributions. *Econometric Theory*, 13(2), 185–213.

Cardell, N. and Dunbar, F. 1980. Measuring the societal impacts of automobile downsizing. *Transportation Research Part A*, 14, 423–34.

Carlsson, F. 2003. The demand for intercity public transport: the case of business passengers. *Applied Economics*, 35, 41–50.

Carrier, E. 2008. *Modeling the Choice of an Airline Itinerary and Fare Product using Booking and Seat Availability Data*. Ph.D. Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Cascetta, E., Nuzzola, A., Russo, F. and Vitetta, A. 1996. A modified logit route choice model overcoming path overlapping problems: Specification and some

calibration results for interurban networks. *Transportation and Traffic Theory,* 697–711.
Castillo, M., Garrow, L. and Lee, M. 2008. *When to Buy a Ticket*. Paper to the Institute for Operations Research and the Management Science Revenue Management and Pricing Section, Montreal, CA.
Cherchi, E., Cirillo, C. and Polak, J.W. 2009. *User Benefits Assessment in Presence of Random Taste Heterogeneity: A Comparison between Parametric and Non-parametric Methods.* Paper to the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.
Cherchi, E. and Ortuzar, J. 2003. *Alternative Specific Variables in Non-linear Utility functions: Influence of Correlation, Homoscedasticity and Taste Variations*. Paper to the Tenth International Conference on Travel Behaviour Research, Lucerne, Switzerland.
Chintagunta, P., Jain, D. and Vilcassim, N. 1991. Investigating heterogeneity in brand preferences in logit models for panel data. *Journal of Marketing Research,* 28(4), 417–28.
Chiou, L. and Walker, J.L. 2007. Masking identification of discrete choice models under simulation methods. *Journal of Econometrics*, 141, 683–703.
Chu, C.A. 1989. *Paired Combinatorial Logit Model for Travel Demand Analysis*. Paper to the Fifth World Conference on Transportation Research: Transport Policy, Management, and Technology Towards 2001, Ventura, CA.
Civil Aeronautics Board. 1970. Effect on total market traffic of changes in quality of service (QSI). Docket 21136 (box 1196). Exhibit BOR-R-300. In Exhibit Series *Rebuttal Exhibits of the Bureau of Operating Rights* by J.F. Adley and C.J. Caridi.
Coldren, G.M. 2005. *Modeling the Competitive Dynamic among Air-travel Itineraries with Generalized Extreme Value Models.* Ph.D. Dissertation, Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL.
Coldren, G.M. and Koppelman, F.S. 2005a. Modeling the competition among air-travel itinerary shares: GEV model development. *Transportation Research Part A,* 39(4), 345–65.
Coldren, G.M. and Koppelman, F.S. 2005b. Modeling the proximate covariance property of air travel itineraries along the time-of-day dimension. *Transportation Research Record,* 1915, 112–23.
Coldren, G.M., Koppelman, F.S., Kasturirangan, K. and Mukherjee, A. 2003. Modeling aggregate air-travel itinerary shares: Logit model development at a major U.S. airline. *Journal of Air Transport Management,* 9(6), 361–69.
Collins, A.T., Rose, J.M. and Hess, S. 2009. *Search Based Internet Surveys: Airline Stated Choice*. Paper to the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.
Cosslett, S. 1981. Maximum likelihood estimation for choice-based samples. *Econometrica,* 49, 1289–1316.

Daganzo, C. 1979. *Multinomial Probit: The Theory and its Applications to Demand Forecasting.* New York: Academic Press.

Dalal, S. and Klein, R. 1988. A flexible class of discrete choice models. *Marketing Science*, 7(3), 232–51.

Daly, A. and Bierlaire, M. 2006. A general and operational representation of generalised extreme value models. *Transportation Research Part B*, 40(4), 285–305.

Data Base Products Inc. 2008. *Superset* [Online]. Available at: http://www.airlinedata.com/products.htm [accessed: 24 September 2008].

de Palma, A., Fontan, C. and Picard, N. 2003. *Departure Time Choice: Estimation Results and Simulation to Paris Area.* Paper to the Tenth International Conference on Travel Behaviour Research, Lucerne, Switzerland.

Dener, A. and Sungur, E. 1991. Some new models for multivariate extremes, in Proceedings of First International Conference on Statistical Computing: The Frontiers of Statistical Computation, Simulation and Modeling: Volume I, edited by P.R. Nelson. Columbus, OH: American Science Press, 185–96.

Dennis, J.E. and Schnabel, R.B. 1996. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Domencich, T.A. and McFadden, D. 1975. *Urban Travel Demand – A Behavioral Analysis*. New York: American Elsevier Publishing Company, Inc.

Dong, X. and Koppelman, F.S. 2003. *Mass Point Mixed Logit: Development and Application.* Paper to the Tenth International Conference on Travel Behaviour Research, Lucerne, Switzerland.

Duarte, A., Garcia, C., Limao, S. and Polydoropoulou, A. 2009. *Experienced and Expected Happiness in Transport Mode Decision Making Process*. Paper to the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

Econometric Software Inc. 2008. *Limdep Version 9* [Online]. Available at: http://www.limdep.com [accessed: 21 September 2008].

Elm-Works Inc. 2008. *ELM Software Version 1.1.10* [Online]. Available at: http://www.elm-works.com [accessed: 21 September 2008].

Eluru, N., Senar, I.N., Bhat, C.R., Pendyala, R.M. and Axhausen, K.W. 2009. *Understanding Residential Mobility: A Joint Model of the Reason for Residential Relocation and Stay Duration*. Paper to the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

Fishman, G. 1996. *Monte Carlo: Concepts, Algorithms, and Applications.* New York: Springer.

Fosgerau, M. 2006. Investigating the distribution of the value of travel time savings. *Transportation Research Part B*, 40(8), 688–707.

Fosgerau, M. and Bierlaire, M. 2007. A practical test of the choice of mixing distributions in discrete choice models. *Transportation Research Part B*, 41(7), 784–94.

Galileo. 2009. *Company Information* [Online]. Available at: http://www.galileo.com/galileo/en-gb [accessed: 2 June 2009].

Gallego, G. and Sahin, O. 2006. Inter-temporal valuations, product design and revenue management. *Working Paper*, Department of Industrial Engineering and Operations Research, Columbia University, New York.

Garrido, R. 2003. *Estimation Performance of Low Discrepancy Sequences in Stated Preference.* Paper to the 10th International Conference on Travel Behaviour Research, Lucerne, Switzerland.

Garrow, L.A. 2009. Online travel data: A goldmine of new opportunities. *Journal of Revenue and Pricing Management* 8(2/3), 247–254.

Garrow, L.A. 2004. Comparison of Choice Models Representing Correlation and Random Taste Variation: An Application to Airline Passengers' Rescheduling Behavior. Ph.D. Dissertation, Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL.

Garrow, L.A. and Bodea, T.D. 2005. *A Rigorous Framework for Empirically Comparing Mixed GEV and Mixed MNL with Error Components Models.* Paper to the European Transport Conference, Strasbourg, France.

Garrow, L.A., Bodea, T.D. and Lee, M. 2009. Generation of synthetic datasets for discrete choice analysis. *Transportation*. Published online on October 14, 2009 and available at http://www.springerlink.com/content/578v28756x36627.

Garrow, L.A., Jones, S.P. and Parker, R.A. 2007. How much airline customers are willing to pay: An analysis of price sensitivity in online distribution channels. *Journal of Revenue and Pricing Management*, 6(1), 1–20.

Garrow, L.A. and Koppelman, F.S. 2004a. Multinomial and nested logit models of airline passengers' no-show and standby behavior. *Journal of Revenue and Pricing Management* 3(3): 237–53.

Garrow, L.A. and Koppelman, F.S. 2004b. Predicting air travelers' no-show and standby behavior using passenger and directional itinerary information. *Journal of Air Transport Management* 10: 401–11.

Gaudry, M. and Dagenais, M. 1978. The Dogit model. *Transportation Research Part B,* 13(2), 105–11.

Goett, A.A., Hudson, K. and Train, K.E. 2000. Customers' choice among retail energy suppliers: The willingness-to-pay for service attributes. *The Energy Journal,* 21(4), 1–28.

Gopinath, D., Schofield, M.L., Walker, J.L. and Ben-Akiva, M. 2005. *Comparative Analysis of Discrete Choice Models with Flexible Substitution Patterns.* Paper to the 84th Annual Meeting of the Transportation Research Board, Washington, D.C.

Greene, W.H. 2007. *Econometric Analysis*. 6th Edition. Upper Saddle River, New Jersey: Prentice Hall.

Greene, W.H. and Hensher, D.A. 2010. *Modeling Ordered Choices: A Primer and Recent Developments*. Cambridge, United Kingdom: Cambridge University Press.

Guevara, C.A., Cherchi, E. and Moreno, M. 2009. *Estimating Random Coefficients Logit Models with Full Covariance Matrix: Comparing the Performance of Mixed Logit and Laplace Approximation Methods.* Paper to the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

Gumbel, E.J. 1935. Les Valeurs Extremes des distribution statistiques. *Annales Henri Poincare*, 4, 115–58.

Gumbel, E.J. 1944. Ranges and midranges. *The Annals of Mathematical Statistics*, 15(4), 414–22.

Gumbel, E.J. 1958. *Statistics of Extremes*. New York: Columbia University Press.

Gumbel, E.J. and Mustafi, C.K. 1967. Some analytical properties of bivariate extreme distributions. *Journal of the American Statistical Association*, 62, 569–88.

Habib, M.A. and Miller, E.J. 2009. *Reference Dependent Residential Location Choice Model within a Relocation Context*. Paper to the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

Hajivassiliou, V. and Ruud, P. 1994. Classical estimation methods for LDV models using simulation, in *Handbook of Econometrics*, *IV*, edited by R. Engle and D. McFadden. New York: Elsevier Science.

Halton, J. 1960. On the efficiency of evaluating certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik,* 2, 84–90.

Han, B., Algers, S. and Engelson, L. 2001. *Accommodating Drivers' Taste Variation and Repeated Choice Correlation in Route Choice Modeling by Using the Mixed Logit Model.* Paper to the 80th meeting of the Transportation Research Board, Washington, D.C.

Harteveldt, H.H., Wilson, C.P., and Johnson, C. 2004. Why leisure travelers book at their favorite sites, in *Forrester Research: Trends* [Online]. Available at: http://www.forrester.com/Research/Document/Excerpt/0,7211,34772,00.html [accessed: 9 May 2009].

Harteveldt, H.H., Johnson, C., Stromberg, C. and Tesch, B. 2006. Leisure travel data overview, in *Forrester Research: Data Overview* [Online]. Available at http://www.forrester.com/Research/Document/Excerpt/0,7211,37715,00.html [accessed: 9 May 2009].

Hensher, D. 1998. Establishing a fare elasticity regime for urban passenger transport. *Journal of Transport Economics and Policy*, 32(2), 221–46.

Hensher, D. 2001a. Measurement of the valuation of travel time savings. *Journal of Transport Economics and Policy*, 35(1), 71–98.

Hensher, D. 2001b. The valuation of commuter travel time savings for car drivers: Evaluating alternative model specifications. *Transportation,* 28(2), 101–18.

Hensher, D. 2001c. Value of passenger travel time savings in air travel: A comprehensive assessment of the published literature. *Technical Memorandum to Boeing Commercial Airplanes*, Seattle, WA.

Hensher, D. 2006. The signs of the time: Imposing a globally signed condition on willingness to pay distributions. *Transportation*, 33(3), 205–22.

Hensher, D. and Greene, W. 2003. The mixed logit model: The state of practice. *Transportation,* 30(2), 133–76.

Hensher, D.A., Greene, W.H., and Rose, J.M. 2005. *Applied Choice Analysis: A Primer*. Cambridge, United Kingdom: Cambridge University Press.

Hensher, D., Louviere, J. and Swait, J. 1999. Combining sources of preference data. *Journal of Econometrics*, 89, 197–221.

Herriges, J. and Kling, C. 1996. Testing the consistency of nested logit models with utility maximization. *Economic Letters*, 50, 33–9.

Hess, S. 2007. *Posterior Analysis of Random Taste Coefficients in Air Travel Choice Behaviour Modeling.* Paper to the 86th Annual Meeting of the Transportation Research Board, Washington, D.C.

Hess, S., Bierlaire, M. and Polak, J.W. 2005a. *Capturing Taste Heterogeneity and Correlation Structure with Mixed GEV Models.* Paper to the 84th Annual Meeting of the Transportation Research Board, Washington, DC.

Hess, S., Bierlaire, M. and Polak, J.W. 2005b. *Discrete Mixture of GEV Models.* Paper to the Fifth Swiss Transport Research Conference, Ascona, Switzerland.

Hess, S. and Polak, J.W. 2005a. Accounting for random taste heterogeneity in airport choice modeling. *Transportation Research Record*, 1915, 36–43.

Hess, S. and Polak, J.W. 2005b. Mixed logit modelling of airport choice in multi-airport regions. *Journal of Air Transport Management*, 11(2), 59–68.

Hess, S., Train, K. and Polak, J. 2004. *On the Use of Randomly Shifted and Shuffled Uniform Vectors in the Estimation of a Mixed Logit Model for Vehicle Choice*. Paper to the 83rd Annual Meeting of the Transportation Research Board, Washington, D.C.

Hess, S., Train, K.E. and Polak, J.W. 2006. On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a mixed logit model for vehicle choice. *Transportation Research Part B*, 40(2): 147–63.

Hoel, P.G., Port, S.C., and Stone, C.J. 1971. *Introduction to Statistical Theory.* Boston: Houghton-Mifflin.

Horowitz, J. 1982. An evaluation of the usefulness of two standard goodness-of-fit indicators for comparing non-nested random utility models. *Transportation Research Record*, 874, 19–25.

Hu, M. and Gallego, G. 2007. Dynamic pricing of perishable assets under competition. Department of Industrial Engineering and Operations Research, Columbia University, New York.

IBM Consulting Services. 2004. *Winning at the Margin: The Impact of Low-Cost Carriers in Asia* [Online]. Available at: http://www-03.ibm.com/industries/travel/us/detail/resource/O367866E52824X62.html [accessed: 8 May 2009].

Iliescu, D.C., Garrow, L.A., and Parker, R.A. 2008. A hazard model of U.S. airline passengers' refund and exchange behavior. *Transportation Research Part B* 42(3): 229–42.

International Air Transport Association. 2009. Billing and Settlement Information [Online]. Available at: http://www.iata.org/ps/financial_services/bsp/index.htm [accessed: 2 June 2009].

Jain, D., Vilcassim, N. and Chintagunta, P. 1994. A random-coefficients logit brand-choice model applied to panel data. *Journal of Business and Economic Statistics*, 12(3), 317–28.

Johnson, N.L., Kotz, S. and Balakrishnan, N. 1995. *Continuous Univariate Distributions, Vol 2*. New York: John Wiley & Sons, Inc.

Kawamura, K. 2000. Perceived value of time for truck operators. *Transportation Research Record*, 1725, 31–36.

Karatzas, J. 2009. *SkyEurope.* Presentation to the AGIFORS Scheduling and Strategic Planning Study Group Meeting, Athens, Greece.

Kim, J.-K., Ulfarsson, G.F., Shankar, V.N. and Mannering, F.L. 2009. *Unobserved Heterogeneity and Heteroskedasticity due to Age when Modeling Pedestrian Injury Severity in Motor Vehicle Crashes.* Paper to the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

Kling, C. and Herriges, J. 1995. An empirical investigation of the consistency of nested logit models with utility maximization. *American Journal of Agricultural Economics*, 77, 875–84.

Koppelman, F.S. 2004. *Course Notes for Travel Demand Analysis*. Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL.

Koppelman, F.S. 2008. Closed-form discrete-choice models, in *Handbook of Transport Modeling*, edited by D.A. Hensher and K.J. Button. 2nd Edition. Amsterdam: Elsevier, 257–77.

Koppelman, F. and Bhat, C. 2006. *A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models*. Report prepared for the U.S. Department of Transportation and Federal Transit Authority. Available at: http://www.civil.northwestern.edu/people/koppelman/PDFs/LM_Draft_060131Final-060630.pdf [accessed: 21 September 2008].

Koppelman, F.S., Coldren, G.C. and Parker, R.A. 2008. Schedule delay impacts on air-travel itinerary demand. *Transportation Research Part B*, 42(3), 263–73.

Koppelman, F.S. and Sethi, V. 2000. Closed-form discrete-choice models, in *Handbook of Transport Modeling*, edited by D.A. Hensher and K.J. Button. Amsterdam: Pergamon, 211–28.

Koppelman, F.S. and Wen, C.-H. 1998a. Alternative nested logit models: Structure, properties and estimation. *Transportation Research Part B*, 32(5), 289–98.

Koppelman, F.S. and Wen, C.-H. 1998b. *Paired Combinatorial Logit: Properties, Estimation and Application*. Paper to the Transportation Research Board, Washington, D.C.

Kotz, S., Balakrishnan, N. and Johnson, N.L. 2000. *Continuous Multivariate Distributions Volume 1*. 2nd edition. New York: John Wiley and Sons.

Law, A.M. and Kelton, W.D. 2000. *Simulation Modeling and Analysis*. 3rd edition. Boston: McGraw Hill.

Lemieux, C., Cieslak, M., and Luttmer, K. 2004. RandQMC. User's Guide. A Package for Quasi-Monte Carlo Methods in C. [Online]. Available at: http://www.math.uwaterloo.ca/~clemieux/randqmc/guide.ps [accessed: 7 May 2009].

Lerman, S. and Manski, C. 1979. Sample design for discrete choice analysis of travel behavior: The state of the art. *Transportation Research*, 13A, 29–44.

Lijesen, M.G. 2006. A mixed logit based valuation of frequency in civil aviation from SP-data. *Transportation Research Part E: Logistics and Transportation Review*, 42(2), 82–94.

Long, J.S. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage Publications, Inc.

Long, J.S. and Freese, J. 2003. *Regression Models for Categorical Dependent Variables Using Stata*. College Station, Texas: Stata Publication.

Louviere, J.J., Hensher, D.A. and Swait, J.D. 2000. *Stated Choice Methods: Analysis and Application*. Cambridge, United Kingdom: Cambridge University Press.

Manski, C. 1977. The structure of random utility models. *Theory and Decision*, 8(3), 229–54.

Manski, C. and Lerman, S. 1977. The estimation of choice probabilities from choice based samples. *Econometrica*, 45, 1977–88.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior, in *Frontiers in Econometrics*, edited by P. Zarembka. New York: Academic Press, 105–42.

McFadden, D. 1975. On independence, structure, and simultaneity in transportation demand analysis. *Working Paper 7511*, Institute of Transportation and Traffic Engineering, University of California, Berkeley, CA.

McFadden, D. 1978. Modeling the choice of residential location, in *Spatial Interaction Theory and Residential Location*, edited by A. Karlqvist, *et al*. Amsterdam: North-Holland, 75–96.

McFadden, D. 2001. Economic choices. *The American Economic Review*, 91(3): 351–78.

McFadden, D. and Train, K. 2000. Mixed MNL models for discrete response. *Journal of Applied Economics,* 15(5), 447–70.

Meloni, I., Bez, M., and Spissu, E. 2009. *An Activity-based Model of Women's Activity-travel Patterns.* Paper to the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

Meloni, I., Portoghese, A., Bez, M. and Spissu, E. 2009. *An Activity-Based Model of the Impact of Physical Activities on the Propensity to Choose a Sustainable Model of Travel to Work or School.* Paper to the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

Metropolitan Transport Commission. 1995. *Metropolitan Transportation Commission airline passenger survey.* Final Report. J.D. Franz Research, Oakland, CA.

Munizaga, M.A. and Alvarez-Daziano, R. 2001. *Mixed Logit vs. Nested Logit and Probit Models.* Paper to the Fifth Tri-Annual Invitational Choice Symposium Workshop: Hybrid Choice Models, Formulation and Practical Issues, Asilomar.

Munizaga, M.A. and Alvarez-Daziano, R. 2002. *Evaluation of Mixed Logit as a Practical Modelling Alternative.* Paper to the European Transport Conference, Cambridge, United Kingdom.

Nako, S.M. 1992. Frequent flyer programs and business travellers: An empirical investigation. *The Logistics and Transportation Review*, 28(4), 395–414.

Newman, J.P. 2008a. *Normalization and Disaggregation of Networked Generalized Extreme Value Models.* Ph.D. Dissertation, Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL.

Newman, J.P. 2008b. Normalization of network generalized extreme value models. *Transportation Research Part B*, 42(10): 958–69.

Nocedal, J. and Wright, S.J. 1999. *Numerical Optimization*. New York: Springer-Verlag.

OAG Worldwide Limited. 2008. *Official Airline Guide* [Online]. Available at: http://www.oag.com [accessed: 30 September 2008].

Parker, R., Lonsdale, R., Glans, T. and Zhang, Z. 2005. *The Naïve Universal Market Simulator*. Presentation to the Advanced Market Techniques Forum of the American Marketing Association, Coeur d' Alene, Idaho.

Pathomsiri, S. and Haghani, A. 2005. Taste variations in airport choice models. *Transportation Research Record,* 1915, 27–35.

PhoCusWright. 2004. *The PhoCusWright Consumer Travel Trends Survey* [Online]. Available at http://store.phocuswright.com/phcotrtrsusi.html.

PhoCusWright. 2008. *The PhoCusWright Consumer Travel Trends Survey* [Online]. Available at http://store.phocuswright.com/phcotrtrsusi.html.

Polt, S. 2002. *Revenue Management Tutorial*. Presentation to the AGIFORS Reservations and Yield Management Study Group Meeting, Berlin, Germany.

Post, D., Mang, S. and Spann, M. 2007. *Pricing of Flexible Products*. Presentation to the National Meeting of the Institute for Operations Research and the Management Sciences, Seattle, WA.

Proussaloglou, K. and Koppelman, F.S. 1999. The choice of air carrier, flight, and fare class. *Journal of Air Transport Management*, 5(4), 193–201.

R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing*. [Online]. Available at: http://www.R-project.org [accessed: 21 September 2008].

Ratliff, R. 2006. *Multi-Flight Demand Untruncation with Recapture*. Presentation to the AGIFORS Revenue Management and Cargo Study Group Meeting, Cancun, Mexico.

Ratliff, R.M., Venkateshwara, R., Narayan, C.P. and Yellepeddi, K. 2008. A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings. *Journal of Revenue and Pricing Management*, 7(2), 53–71.

Recker, W.W. 1995. Discrete choice model with an oddball alternative. *Transportation Research Part B*, 29(3), 201–12.

Research and Innovative Technology Administration, Bureau of Transportation Statistics, U.S. Department of Transportation. 2009. *Airline On-time Statistics* [Online]. Available at: http://www.bts.gov/xml/ontimesummarystatistics/src/index.xml [accessed: 2 June 2009].

Revelt, D. and Train, K. 1998. Mixed logit with repeated choices: Households' choices of appliance efficiency level. *Review of Economics and Statistics*, 80(4), 647–57.

Revelt, D. and Train, K. 1999. Customer-specific taste parameters and mixed logit. *Working Paper*, Department of Economics, University of California, Berkeley.

Rouwendahl, J. and Meijer, E. 2001. Preferences for housing, job, and commuting: A mixed logit analysis. *Journal of Regional Science*, 41, 475–505.

Ruud, P.A. 2000. *An Introduction to Classical Econometric Theory*. Oxford: Oxford University Press.

Sabre. 2009. *Company Information* [Online]. Available at: http://www.sabre.com/ [accessed: 2 June 2009].

Sandor, Z. and Train, K. 2004. Quasi-random simulation of discrete choice models. *Transportation Research Part B*, 38(4): 313–27.

SAS. 2008. *SAS Version 9.2* [Online]. Available at: http://www.sas.com [accessed: 21 September 2008].

Sener, I.N., Eluru, N. and Bhat, C.R. 2009. *An Analysis of Bicyclists and Bicycling Characteristics: Who, Why, and How Much are They Bicycling?* Paper to the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

Sethi, V. 2000. *Modeling Complex Substitution Patterns with Variance and Covariance Heterogeneity in Long Distance Travel Choice Models.* Ph.D. Dissertation, Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL.

Shi, D. and Zhou, S. 1999. Moment estimation for multivariate extreme value distribution in a nested logistic model. *Annuals of the Institute of Statistical Mathematics*, 51(2), 253–64.

Shiau, C.-S., Michalek, J.J. and Hendrickson, C.T. 2009. *A Structural Analysis of Vehicle Design Responses to Corporate Average Fuel Economy Policy*. Paper to the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

Small, K.A. 1987. A discrete choice model for ordered alternatives. *Econometrica*, 55(2), 409–57.

Small, K., Winston, C. and Yan, J. 2005. Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica*, 73(4), 1367–82.

Srinivasan, S., Bhat, C.R. and Holguin-Veras, J. Empirical analysis of the impact of security perception on intercity mode choice. *Transportation Research Record*, 1942, 9–15.

StataCorp. 2008. *STATA Version 10* [Online]. Available at: http://www.stata.com [accessed: 21 September 2008].

Stephenson, A. 2003. Simulating multivariate extreme value distributions of logistic type. *Extremes*, 6, 49–59.

Subramanian, S. 2006. *Nested Logit Models for Joint Mode and Destination Choice*. M.S. Thesis, Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL.

Suzuki, Y., Tyworth, J.E. and Novack, R.A. 2001. Airline market share and customer service quality: A reference-dependent model. *Transportation Research Part A*, 35(9), 773–88.

Swait, J. 2000. Choice set generation within the Generalized Extreme Value family of discrete choice models. *Working Paper*, University of Florida, Gainesville, FL.

Swait, J. 2009. Choice models based on mixed discrete/continuous PDFs. *Transportation Research Part B*. 43(7), 766–83.

Swait, J. and Adamowicz, W. 1996. The effect of choice environment and task demands on consumer behavior: Discriminating between contribution and confusion. *Working Paper*, Department of Rural Economy, University of Alberta, Canada.

Swait, J. and Ben-Akiva, M. 1987. Empirical test of a constrained choice discrete model: Mode choice in Sao Paulo, Brazil. *Transportation Research Part B,* 21(2), 103–16.

Swait, J. and Stacey, E.C. 1996. *Consumer Brand Assessment and Assessment Confidence in Models of Longitudinal Choice Behavior*. Presentation to the INFORMS Marketing Science Conference, Gainesville, FL, March 7–10.

Talluri, K.T. and van Ryzin, G. 2004. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50(1): 15–33.

Tiago de Oliveira, J. 1974. Regression in the nondifferentiable bivariate extreme models. *Journal of the American Statistical Association,* 69(347), 816–18.

Tiago de Oliveira, J. 1982. Decision and modelling for extremes, in *Some Recent Advances in Statistics* edited by J. Tiago de Oliveira and B. Epstein. London: Academic Press, 101–10.

Train, K. 1986. *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*. 1st Edition. Cambridge, Massachusetts: The MIT Press.

Train, K. 1993. *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*. 3rd Edition. Cambridge, Massachusetts: The MIT Press.

Train, K. 1998. Recreation demand models with taste variations. *Land Economics,* 74, 230–39.

Train, K. 2000. Halton sequences for mixed logit. *Working Paper*, Department of Economics, University of California, Berkeley.

Train, K.E. 2003. *Discrete Choice Methods with Simulation.* Cambridge, UK: University Press.

Train, K., McFadden, D. and Ben-Akiva, M. 1987a. The demand for local telephone service: A fully discrete model of residential calling patterns and service choice. *Rand Journal of Economics*, 18, 109–23.

Train, K., McFadden, D. and Ben-Akiva, M. 1987b. The goods-leisure tradeoff and disaggregate work trip mode choice models. *Transportation Research*, 12, 349–53.

Train, K. and Sonnier, G. 2005. Mixed logit with bounded distributions of correlated partworths, in *Applications of Simulation Methods in Environmental Resource Economics,* edited by A. Alberini and R. Scarpa. The Netherlands: Springer. 117–34.

Transportation Research Board. 2002. Transportation Research E-Circular E-C040. Aviation Demand Forecasting: A Survey of Methodologies. Washington, D.C.: Transportation Research Board.

van Bladel, K., Bellemans, T., Janssens, D. and Wets, G. 2009. *Activity-travel Planning and Rescheduling Behavior: An Empirical Analysis of Influencing Factors*. Paper to the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

van Ryzin, G. and Liu, Q. 2004. On the choice-based linear programming model for network revenue management. *Decision, Risk & Operations, Working Paper Series DRO-2004-04*, Columbia Business School, New York.

van Ryzin, G. and Vulcano, G. 2008a. Computing virtual nesting controls for network revenue management under customer choice behavior. *Manufacturing and Service Operations Management*, 10(3): 448–67.

van Ryzin, G. and Vulcano, G. 2008b. Simulation-based optimization of virtual nesting controls for network revenue management. *Operations Research*, 56(4): 865–80.

Vulcano, G., van Ryzin, G. and Chaar, W. 2006. *Choice-based Revenue Management: An Empirical Study of Estimation and Optimization*. Presentation to the Annual Meeting of INFORMS, Pittsburgh.

Vovsha, P. 1997. The cross-nested logit model: Application to mode choice in the Tel-Aviv metropolitan area. *Transportation Research Record*, 1607, 6–15.

Walker, J.L. 2001. *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*. Ph.D. Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Walker, J.L. 2002. The logit kernel (or continuous mixed logit) model: Dispelling misconceptions of identification. *Transportation Research* Record, 1805, 86–98.

Walker, J.L., Ben-Akiva, M. and Bolduc, D. 2007. Identification of parameters in normal error component logit-mixture (NECLM) models. *Journal of Applied Econometrics*, 22: 1095–1125.

Warburg, V., Bhat, C. and Adler. T. 2006. Modeling demographic and unobserved heterogeneity in air passengers' sensitivity to service attributes in itinerary choice. *Transportation Research Record*, 1951, 7–16.

Wen, C.-H., Chen, T.-N. and Huang, W.-W. 2009. *Mixed Logit Analysis of International Airline Choice: An Empirical Study of Taiwan's Air Routes*. Paper to the 88th meeting of the Transportation Research Board, Washington, D.C.

Wen, C.-H. and Koppelman, F. 2001. The generalized nested logit model. *Transportation Research Part B*, 35(7), 627–41.

Williams, H. 1977. On the formulation of travel demand models and economic evaluation measures of user benefits. *Environment and Planning A*, 9, 285–344.

Williams, H.C. and Ortuzar, J.D. 1982. Behavioural theories of dispersion and the mis-specification of travel demand models. *Transportation Research Part B*, 16(3), 167–219.

Worldspan. 2009. *Company Information* [Online]. Available at: http://www.worldspan.com/home.asp?fPageID=2 [accessed: 2 June 2009].

Xu, N., Holguin-Veras, J. and Bhat, C. 2009. *An Assessment of the Impacts of Inspection Time on the Airline Industry's Market Share after 9/11.* Paper to the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

Yang, C.-W. and Sung, Y.-C. 2010. Constructing a mixed-logit model with market positioning to analyze the effects of new mode introduction. *Journal of Transport Geography.* 18(1), 175–182.

Zhang, D. and Cooper, W. 2005. Revenue management for parallel flights with customer-choice behavior. *Operations Research*, 53(3): 415–31.

Index

Author Index